Assignment 3

Filip Wilhelm Sjostrand

2023-05-21

```
library(RSQLite)
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
path <- file.path("/Users/filipsjostrand/Documents/UC Davis/Courses/STA 141B/data/STA141B Spring 2023/s
con <- dbConnect(SQLite(), path)
```

1 How many users are there?

For this question, we only need the `Users` table. In the Stack Exchange data schema, each row in the `Users` table represents a unique user. The primary key for the Users table is `Id`, which is unique for each user. We use `DISTINCT()` to ensure no duplicate counts. From the query below we can observe that there are 321677 users on the forum.

```
SELECT COUNT(DISTINCT Id) AS TotalUniqueUsers
FROM Users;
```

Table 1: 1 records

TotalUniqueUsers
321677

2 How many users joined since 2020?

We shall remain using the `Users` table. Within the table we find `CreationDate`. We will extract the year as instructed by Ian (2021) and convert from datetime into integer and look for years greater or equal to 2020. From below we see that since 2020, 100796 users where created.

```
SELECT COUNT(*) AS UsersSince2020
FROM Users
WHERE CAST(strftime('%Y', CreationDate) AS INTEGER) >= 2020;
```

Table 2: 1 records

UsersSince2020
100796

3 How many users joined each year?

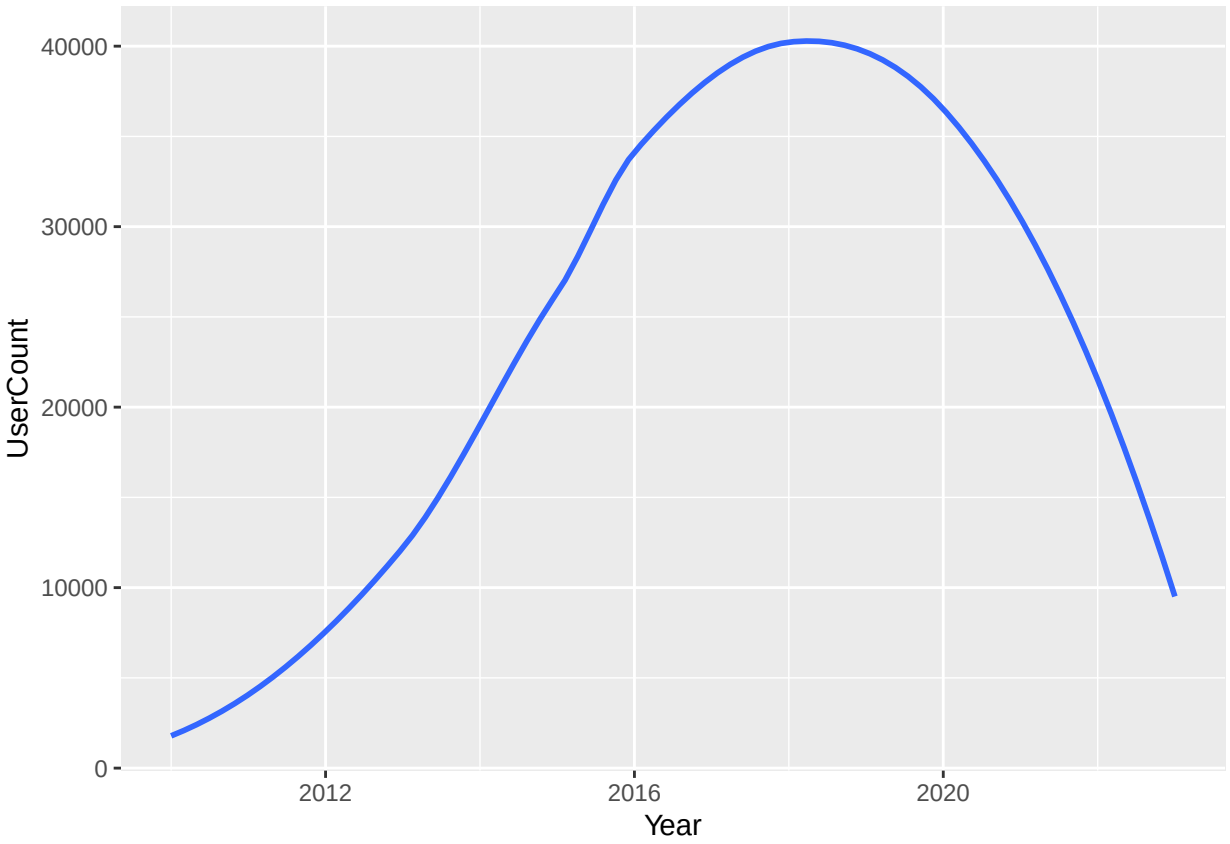
We will remain using the `Users` table. Using similar method as above to extract the year, we then group by each year to create a data frame with the count of each user per year.

```
SELECT
  CAST(strftime('%Y', CreationDate) AS INTEGER) AS Year,
  COUNT(*) AS UserCount
FROM Users
GROUP BY Year
ORDER BY Year;
```

The observed trend of increasing new user registrations on Stack Overflow until 2018, followed by a subsequent decrease, could be attributed to various factors. These include the platform's growing popularity and eventual saturation, competition from other platforms, and potential changes in platform policies or features.

```
ggplot(df, aes(x = Year, y = UserCount)) +
    geom_smooth(se = FALSE)
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



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

We start with the `Posts` table. Types are referred to by IDs, which are linked to `value` in the `PostTypeIdMap` table (Community Wiki, 2023). We join these tables on this ID, group by the `value`, count the posts, and sort in descending order to get a more general table used to answer several questions. First, There appears to be 8 different types of posts.

```sql
SELECT ptm.value, COUNT(*) AS PostCount
FROM Posts p
INNER JOIN PostTypeIdMap ptm ON p.PostTypeId = ptm.Id
GROUP BY ptm.value
ORDER BY PostCount DESC;
```

Table 3: 7 records

value	PostCount
Question	204370
Answer	197928
Tag wiki excerpt	1444
Tag wiki	1444
Moderator nomination	23
Orphaned tag wiki	6
Wiki placeholder (seems to only be the election description)	5

5 How many posted questions are there?

From Table 3 above, we can derive that there are 204370 questions.

6 What are the top 50 most common tags on questions? For each of the top 50 tags on questions, how many questions are there for each tag?

The provided SQL query uses the `TagPosts` table. By comparing the ID of `TagPosts` and `Posts` we can confirm that they are referring to the same. Hence, we can count the unique occurrences in `TagPosts`. The query groups rows by tag, counts the number of occurrences (the number of questions with that tag), and orders the result in descending order of count, limiting the output to the top 50 tags. Below we can observe the different tags and in how many questions they occur. *r*, *regression*, and *machine-learning* dominate.

```
SELECT Tag, COUNT(*) AS CountTags
FROM TagPosts
GROUP BY Tag
ORDER BY CountTags DESC
LIMIT 50;
```

Table 4: Displaying records 1 - 10

Tag	CountTags
r	28495
regression	28146
machine-learning	19355
time-series	13745
probability	11894
hypothesis-testing	10091
distributions	9147
self-study	7985
neural-networks	7793
bayesian	7628

7 How many tags are in most questions?

We shall count the average number of tags per question to estimate this. By simply counting how many times each `Id` appears in `TagPosts` we know how many tags a question have. THen we find the average. It appears as if 3 tags per question is the most common occurance.

```
SELECT AVG(NumTag)
FROM (SELECT COUNT(Id) AS NumTag FROM TagPosts GROUP BY Id);
```

Table 5: 1 records

AVG(NumTag)
3.08534

8 How many answers are there?

As observable in Table 3, there are 197928 answers.

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

In the `Posts` table, each post has a `PostTypeId` (where we know from before that 1 refers to questions (Community Wiki, 2023)) and a `CreationDate`. The query filters for questions and then sorts them by `CreationDate` in descending order. The `LIMIT 1` clause restricts the result to the top row. Hence, the latest post from our database was on 2023-03-05.

9.1 Find it on the stats.exchange.com Web site and provide the URL

Here is the link to that post: https://stats.stackexchange.com/questions/608458/are-there-any-methods-that-combine-mcmc-and-vi

9.2 How would we map a question in the Posts table to the corresponding SO URL?

It appears as if the question ID is mentioned in the URL, which should be a good way to map the URL to the table.

```
SELECT CreationDate, Id
FROM Posts
WHERE PostTypeId = 1
ORDER BY CreationDate DESC
LIMIT 1;
```

Table 6: 1 records

CreationDate	Id
2023-03-05T05:10:18.393	608405

10 For the 10 users who posted the most questions

- How many questions did they post?

- What are the users' names?

- When did they join SO?

- What is their Reputation?

- What country do they have in their profile?

The `Users` table contains user details such as `DisplayName`, `CreationDate`, `Reputation`, and `Location`, and the `Posts` table contains posts details with `OwnerUserId` referring back to the `Users` table and `PostTypeId` distinguishing questions. The query joins these two tables on user ID, filters for questions, groups by user, and orders by the count of questions in descending order, limiting to the top 10 users.

stats_noob seems to be an avid poster of questions. However, it is slightly worrisome that he has such a poor reputation. Either the data base is wrong or he posts rather poor questions.

```
SELECT
  u.Id AS UserId,
  u.DisplayName AS UserName,
  u.CreationDate AS JoinDate,
  u.Reputation AS Reputation,
  u.Location AS Country,
  COUNT(p.Id) AS QuestionCount
FROM Users u
INNER JOIN Posts p ON u.Id = p.OwnerUserId
WHERE p.PostTypeId = 1
GROUP BY UserId
ORDER BY QuestionCount DESC
LIMIT 10;
```

Table 7: Displaying records 1 - 10

UserId	UserName	JoinDate	Reputation	Country	QuestionCount
77179	stats_noob	2015-05-14T21:12:31.790	1		349
1005	Tim	2010-08-19T15:31:09.537	18497		298
9162	user1205901 -	2012-02-13T02:09:08.377	11859		264
53690	Richard Hardy	2014-08-08T10:57:13.613	60742	Europe	255
108150	user321627	2016-03-10T14:45:28.010	2478		236
113777	Haitao Du	2016-04-27T20:51:38.203	34665		192
28986	Charlie Parker	2013-08-09T19:20:37.540	6286		184
40252	An old man in the sea.	2014-02-14T13:18:39.917	5330		180
163242	The Pointer	2017-05-30T00:13:41.380	1344		166
56211	rnso	2014-09-22T08:35:18.697	9299		164

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?

The `Badges` table has a `Class` field, where '1' represents gold, '2' represents silver, and '3' represents bronze (Community Wiki, 2023). The query first identifies the top 10 users who posted the most questions (as in the previous question) and then joins these users with their badges, finally using the method, as presented by BMN (2013), of counting the badges of each class.

Here, the user *Tim* is a leader in all three classes. We can also observe that *stats_noob* dropped on the leader board, somewhat instigating what I previously mentioned (poor questions)

```
WITH TopUsers AS (
  SELECT
    u.Id AS UserId,
    u.DisplayName AS UserName,
    u.CreationDate AS JoinDate,
    u.Reputation AS Reputation,
    u.Location AS Country,
    COUNT(p.Id) AS QuestionCount
  FROM Users u
  INNER JOIN Posts p ON u.Id = p.OwnerUserId
  WHERE p.PostTypeId = 1
  GROUP BY UserId
  ORDER BY QuestionCount DESC
  LIMIT 10
  )

SELECT
  tu.UserId,
  u.DisplayName AS UserName,
  COUNT(CASE WHEN b.Class = 1 THEN 1 END) AS GoldBadges,
  COUNT(CASE WHEN b.Class = 2 THEN 1 END) AS SilverBadges,
  COUNT(CASE WHEN b.Class = 3 THEN 1 END) AS BronzeBadges
FROM TopUsers tu
INNER JOIN Users u ON tu.UserId = u.Id
INNER JOIN Badges b ON tu.UserId = b.UserId
GROUP BY tu.UserId
ORDER BY
  GoldBadges DESC,
  SilverBadges DESC,
  BronzeBadges DESC;
```

Table 8: Displaying records 1 - 10

UserId	UserName	GoldBadges	SilverBadges	BronzeBadges
1005	Tim	32	116	205
9162	user1205901 -	26	88	161
113777	Haitao Du	19	130	228
56211	rnso	18	59	109
28986	Charlie Parker	13	65	122
53690	Richard Hardy	12	114	237

UserId	UserName	GoldBadges	SilverBadges	BronzeBadges
40252	An old man in the sea.	4	27	65
108150	user321627	4	20	63
77179	stats_noob	2	32	68
163242	The Pointer	0	22	44

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

We wish to first find all the questions. As known, we filter on `PostTyoeId = 1`. Then, the query searches for each term within the `Body` field of questions in the `Posts` table. It utilizes the same method given by BMN (2013).

Given that we already know that both Regression and Machine Learning where in the top 3 of tags it is to no surprise we also see them at the top in in-text mentions. However, interseting that ANOVA has more than machine learning as it is not even in the top 10 of tags.

```
SELECT
  SUM(CASE WHEN Body LIKE '%Regression%' THEN 1 ELSE 0 END) AS Regression,
  SUM(CASE WHEN Body LIKE '%ANOVA%' THEN 1 ELSE 0 END) AS ANOVA,
  SUM(CASE WHEN Body LIKE '%Data Mining%' THEN 1 ELSE 0 END) AS DataMining,
  SUM(CASE WHEN Body LIKE '%Machine Learning%' THEN 1 ELSE 0 END) AS MachineLearning,
  SUM(CASE WHEN Body LIKE '%Deep Learning%' THEN 1 ELSE 0 END) AS DeepLearning,
  SUM(CASE WHEN Body LIKE '%Neural Network%' THEN 1 ELSE 0 END) AS NeuralNetwork
FROM Posts
WHERE PostTypeId = 1;
```

Table 9: 1 records

Regression	ANOVA	DataMining	MachineLearning	DeepLearning	NeuralNetwork
39296	7167	405	5404	1124	5063

13 Using the Posts and PostLinks tables, how many questions gave rise to a ”related” or ”duplicate” question?

13.1 - And how many responses did these questions get?

13.2 - How experienced were the users posting these questions.

First, we find out how many questions gave rise to a “related” or “duplicate” question. In the `PostLinks` table, we’ll consider `LinkTypeId` values of 3, which represent “duplicate” links (Community Wiki, 2023). We’ll count the distinct `PostId` values, which represent the original questions that have related or duplicate links. Hence 12326 occurrences of duplicate posts

```sql
SELECT COUNT(DISTINCT(PostId)) AS NumDuplicate
FROM PostLinks
WHERE LinkTypeId = 3;
```

Table 10: 1 records

NumDuplicate
12326

Next, we find out how many responses these questions got. We'll count the number of posts in the `Posts` table where `ParentId` is one of the `PostId` values identified above. `PostTypeId` is 2, representing answers (Community Wiki, 2023). So, 4136 of the duplicate questions got responses

```sql
WITH Duplicate AS (
  SELECT DISTINCT PostId
  FROM PostLinks
  WHERE LinkTypeId = 3
)

SELECT COUNT(*) AS NumResponses
FROM Posts
WHERE
  ParentId IN (SELECT PostId FROM Duplicate) AND
  PostTypeId = 2;
```

Table 11: 1 records

NumResponses
4136

Lastly, we shall find out how experienced the users posting these questions were. We'll join the `Posts` and `Users` tables to get the `Reputation` of the users who posted the original questions identified above. However, we will want to find the latest of all posts since a users reputation may have increased or decreased. We'll then calculate a summary statistic: the average.

In this query, `MAX(p.CreationDate)` is used to select the latest `CreationDate` for each unique `OwnerUserId`. The `GROUP BY` statement is necessary to specify that we wish the max of each user, in this case, we are grouping by `OwnerUserId`.

We see that the average experience is around 475, which implies that they are not complete beginners who post duplicate questions. However, as apparent in the tables above: you should only expect an answer one third of the time.

```sql
WITH Duplicate AS (
  SELECT DISTINCT(PostId)
  FROM PostLinks
  WHERE LinkTypeId = 3
),

LatestReputation AS (
```

```
  SELECT
    p.OwnerUserId,
    u.Reputation,
    MAX(p.CreationDate) AS LatestCreationDate
  FROM Posts p
  INNER JOIN Users u ON p.OwnerUserId = u.Id
  WHERE p.Id IN (SELECT PostId FROM Duplicate)
  GROUP BY p.OwnerUserId
)

SELECT AVG(Reputation) AS MeanExperience FROM LatestReputation;
```

Table 12: 1 records

MeanExperience
475.8029

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

The `Posts` table has a `CreationDate` field which records the date each post was created, and a `PostTypeId` field which distinguishes between questions and answers. The query groups by `PostTypeId`, filters for questions and answers, and finds the minimum and maximum `CreationDate` for each group.

We can observe that both questions and answer begun 2009-02-02 and kept going til today (or the day of the last update of the data base). Suggest an active community.

```
SELECT
  ptm.value AS Kind,
  MIN(CreationDate) AS EarliestDate,
  MAX(CreationDate) AS LatestDate
FROM Posts p
INNER JOIN PostTypeIdMap ptm ON p.PostTypeId = ptm.id
WHERE PostTypeId IN (1, 2)
GROUP BY PostTypeId;
```

Table 13: 2 records

Kind	EarliestDate	LatestDate
Question	2009-02-02T14:21:12.103	2023-03-05T05:10:18.393
Answer	2009-02-02T14:24:31.740	2023-03-05T04:48:34.853

15 What question has the most comments associated with it?

The `Comments` table has a `PostId` field which links each comment to a post. By grouping comments by `PostId`, we can count the number of comments for each question. Then we find the `PostId` with the highest comment count. Finally, we count the number of posts where this `PostId` is the `ParentId` and `PostTypeId`

is 2, representing answers. However, we shall also consider that it have to be a question and not an answer, thus, looking at `PostTypeId` as 1, too.

We find that the most comments associate with a question is 54. Given the nature of the forum it is quite a high count since most questions and answers—from experience—rarely have 3 or more comments to them.

```
SELECT c.PostId, COUNT(*) AS CommentCount
FROM Comments c
INNER JOIN Posts p ON c.PostId = p.Id
WHERE PostTypeId = 1
GROUP BY PostId
ORDER BY CommentCount DESC
LIMIT 1;
```

Table 14: 1 records

PostId	CommentCount
328630	54

15.1 How many answers are there for this question?

Given what i mentioned above, the number 54 is even considered more extreme now that only 6 answers were given to this question. From a quick look, the questions regards the appropriateness of ridge regressions in higher dimensions.

```
WITH Answers AS (
  SELECT *
  FROM Posts p
  WHERE p.PostTypeId = 2
    AND ParentId = 328630
  )

SELECT ParentId, COUNT(*) AS NumAnswers
FROM Answers;
```

Table 15: 1 records

ParentId	NumAnswers
328630	6

16 How many comments are there across all posts?

- How many posts have a comment?

The first two questions can be answered directly from the `Comments` table. All comments is simply all the rows of the table, while posts with comments are all the unique posts IDs in the table. For the third question, we shall find how many comments each question has.

It appears as if there are 768069 comments in total but only 229859 posts who have comments.

```sql
SELECT
  COUNT(*) AS TotalComments,
  COUNT(DISTINCT PostId) AS PostsWithComments
FROM Comments;
```

Table 16: 1 records

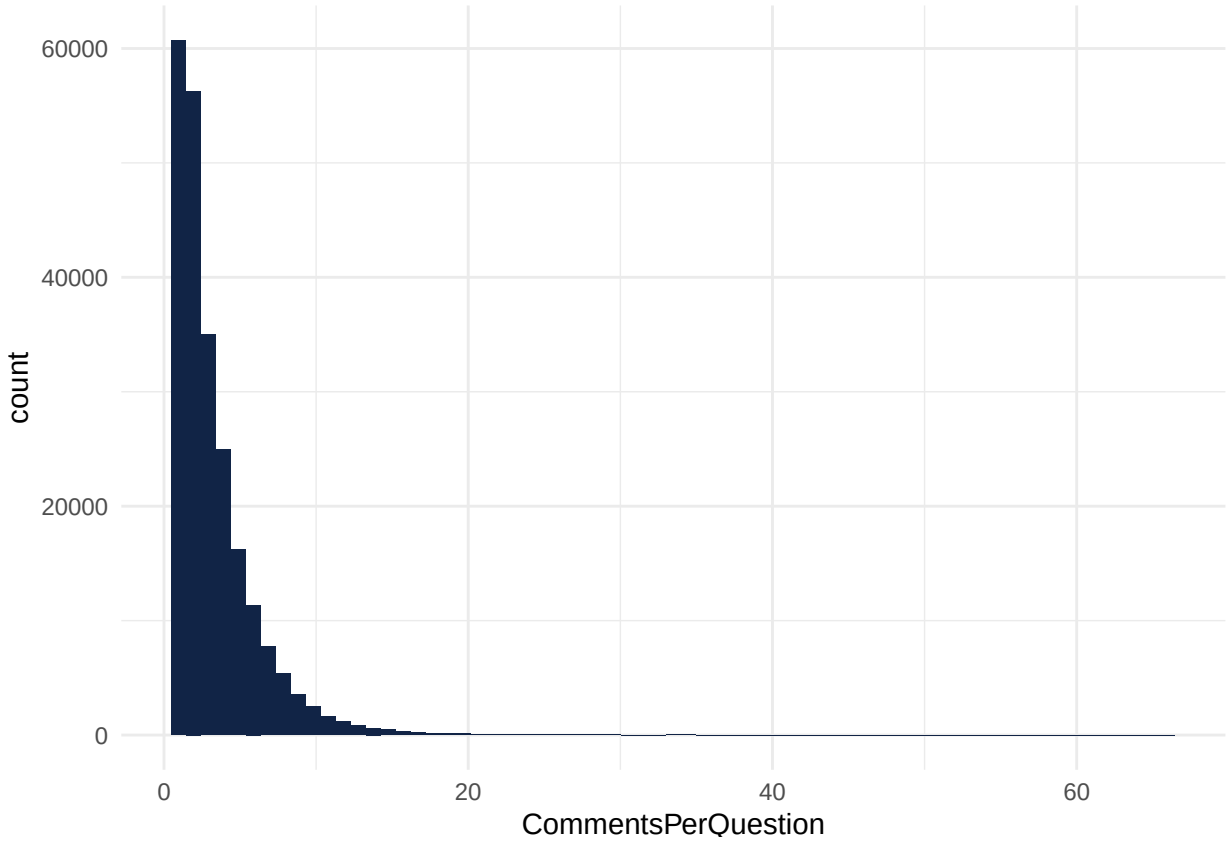
TotalComments	PostsWithComments
768069	229859

16.1 What is the distribution of comments per question?

We see a negative exponential distribution

```sql
SELECT PostId, COUNT(*) AS CommentsPerQuestion
FROM Comments
GROUP BY PostId
```

```
ggplot(df) +
  aes(x = CommentsPerQuestion) +
  geom_histogram(bins = 67L, fill = "#112446") +
  theme_minimal()
```



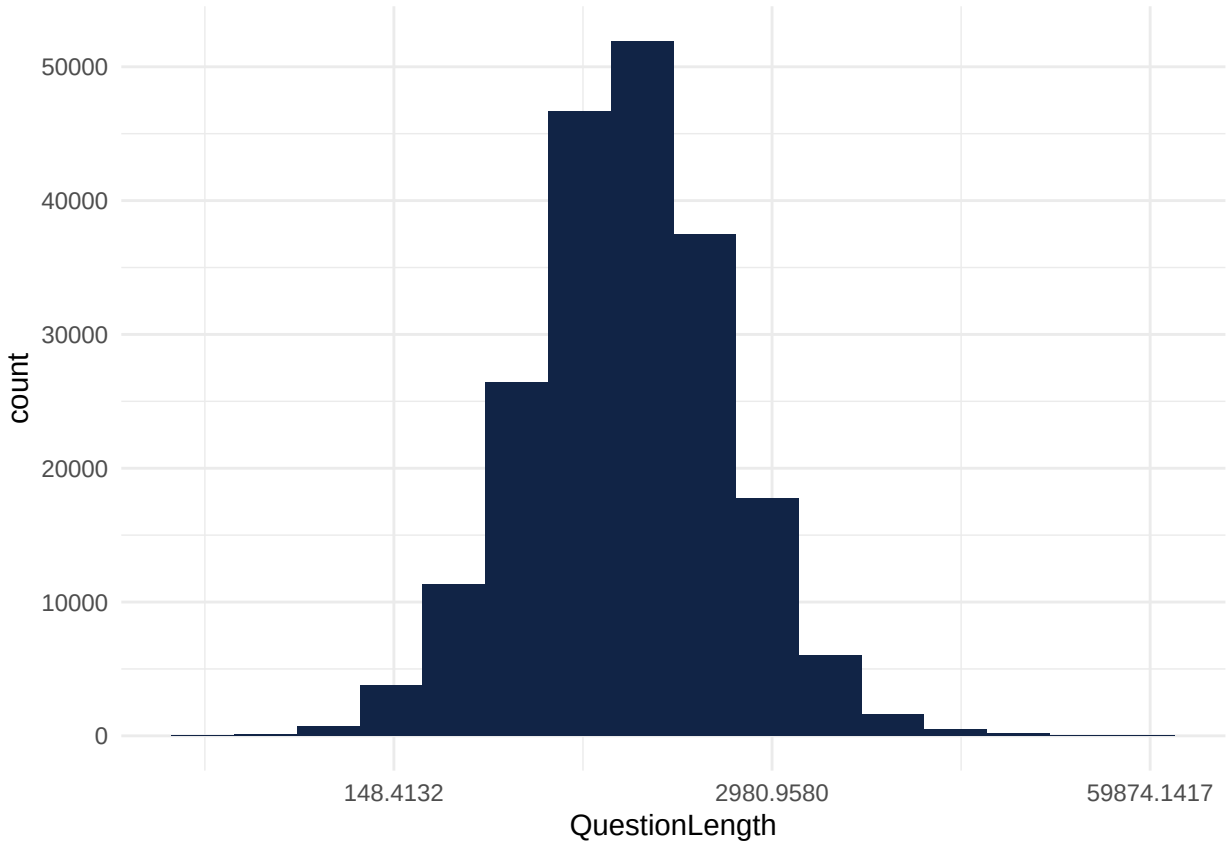
17 Is there any relationship between the number of tags on a question, the length of the question, and the number of responses (posts and comments)?

To count the number of tags on each question we utilize the method suggested by b8b8j (2010). Note that the tags are a string with the format …, so we can count tags by counting the number of < characters. The length of the questions are straight forward found by `LEN`. For the number of response we count the rows for which `PostTypeId` is either 2 (answer) or 3 (comment). Finally we join the tables together.

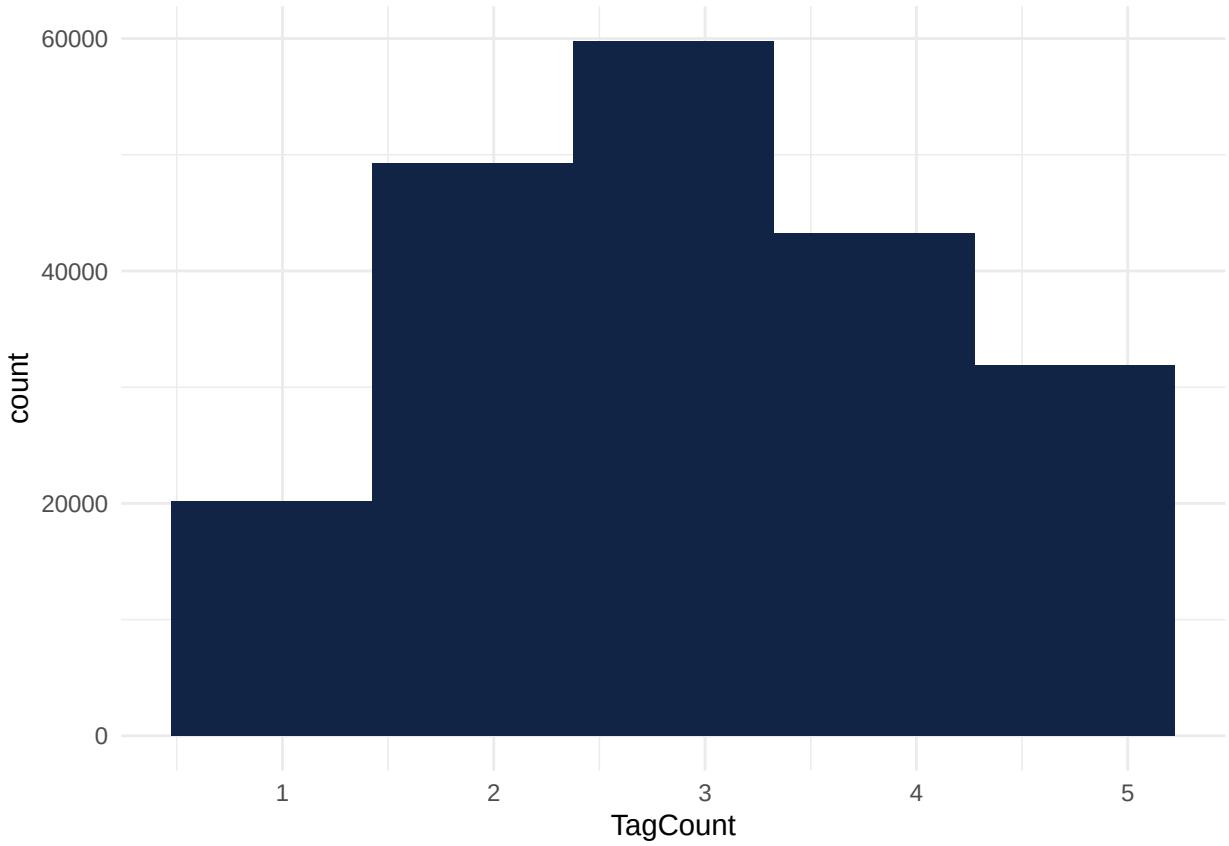
We begun by checking the normality of each population. We found that the question length turned normal using a log transformation, tag count was approximately normally distributed, but no success in transforming response count as it was extremely centered around 1. So, we utilized a persons correlation (parametric) between tag count and log-transformed question length. Between response count and the others we used Spearmans Rho (non-parametric). In conclusion we see a significant relationship in each instance. However, the strength in the association of response rate and the others is extremely low. Between tag count and questions lenght there is a weak associaton. Hence we could conclude some weak association in an increase in question lenght and the number of tags—which seems reasonable.

```
WITH QuestionTags AS (
  SELECT
    Id,
    LENGTH(Tags) - LENGTH(REPLACE(Tags, '<', '')) AS TagCount,
    LENGTH(Body) AS QuestionLength
  FROM Posts
  WHERE PostTypeId = 1
),

QuestionResponses AS (
  SELECT ParentId, COUNT(*) AS ResponseCount
  FROM Posts
  WHERE PostTypeId IN (2, 3)
  GROUP BY ParentId
)

SELECT
  qt.Id,
  qt.TagCount,
  qt.QuestionLength,
  qr.ResponseCount
FROM QuestionTags qt
LEFT JOIN QuestionResponses qr ON qt.Id = qr.ParentId;
```

```
ggplot(df) +
  aes(x = QuestionLength) +
  geom_histogram( bins = 16L, fill = "#112446") +
  scale_x_continuous(trans = "log") +
  theme_minimal()
```

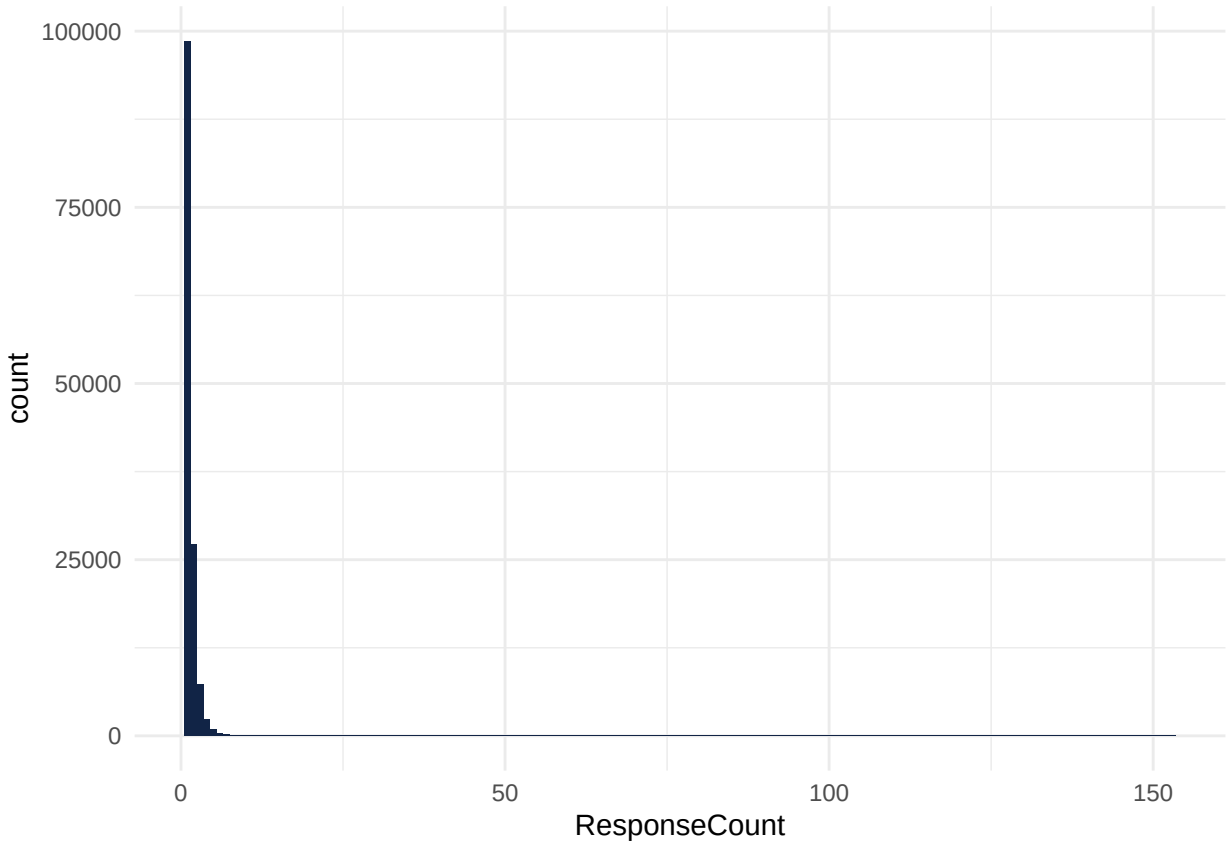



```
ggplot(df) +
  aes(x = TagCount) +
  geom_histogram(bins = 10L, fill = "#112446", binwidth = .95) +
  theme_minimal()
```




```
ggplot(df) +
  aes(x = ResponseCount) +
  geom_histogram(bins = 10L, fill = "#112446", binwidth = 1) +
  theme_minimal()
```

```
## Warning: Removed 66970 rows containing non-finite values (`stat_bin()`).
```




```
cor.test(df$TagCount, log(df$QuestionLength))
```

```
##
##  Pearson's product-moment correlation
##
## data:  df$TagCount and log(df$QuestionLength)
## t = 100.84, df = 204368, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2135679 0.2218280
## sample estimates:
##       cor
## 0.2177019
```

```
cor.test(df$TagCount, df$ResponseCount, method = 'spearman')
```

```
## Warning in cor.test.default(df$TagCount, df$ResponseCount, method =
## "spearman"): Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  df$TagCount and df$ResponseCount
## S = 4.3694e+14, p-value = 7.581e-05
## alternative hypothesis: true rho is not equal to 0
```

```
## sample estimates:
##         rho
## -0.01067564
```

```
cor.test(log(df$QuestionLength), df$ResponseCount, method = 'spearman')
```

```
## Warning in cor.test.default(log(df$QuestionLength), df$ResponseCount, method =
## "spearman"): Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  log(df$QuestionLength) and df$ResponseCount
## S = 4.4767e+14, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##         rho
## -0.03549252
```

18 Do the people who vote tend to have badges?

To answer this, we first identify users who have voted using the `Votes` table. Then we identify users who have badges using the `Badges` table. We compare these two sets of users to find users who have voted and have badges, and users who have voted but do not have badges. We observe that a great majority of voters have badges.

```
WITH Voted AS (
  SELECT UserId
  FROM Votes
  GROUP BY UserId
),

WithBadges AS (
  SELECT UserId
  FROM Badges
  GROUP BY UserId
)

SELECT
  COUNT(DISTINCT v.UserId) AS Voters,
  COUNT(DISTINCT wb.UserId) AS VotersWithBadges,
  (COUNT(DISTINCT v.UserId) - COUNT(DISTINCT wb.UserId)) AS VotersWithoutBadges
FROM Voted v
LEFT JOIN WithBadges wb ON v.UserId = wb.UserId;
```

Table 17: 1 records

Voters	VotersWithBadges	VotersWithoutBadges
3191	3186	5

19 How many questions were edited by the original poster? by other users?

Edits are recorded in the `PostHistory` table, and we can assume that an 'edit' as stated in the question corresponds to a `PostHistoryTypeId` of 4 (Edit Tags), 5 (Edit Title), or 6 (Edit Body) (Community Wiki, 2023). We can link the `PostHistory` entries to the original posts using the `PostId` field, and determine whether the edit was made by the original poster or by another user by comparing the `UserId` field in `PostHistory` with the `OwnerUserId` field in `Posts`. We see that most of the time, questions are edited by its original creator.

```
SELECT
  CASE
    WHEN ph.UserId = p.OwnerUserId THEN 'Original'
    ELSE 'Others'
  END AS Editor,
  COUNT(*) AS Count
FROM PostHistory ph
INNER JOIN Posts p ON ph.PostId = p.Id
WHERE ph.PostHistoryTypeId IN (4, 5, 6)
GROUP BY Editor;
```

Table 18: 2 records

Editor	Count
Original	269921
Others	183297

20 How many posts have multiple different people who edit it?

Edits are recorded in the `PostHistory` table, and we can assume, again, that an edit corresponds to a `PostHistoryTypeId` of 4 (Edit Tags), 5 (Edit Title), or 6 (Edit Body). By grouping the `PostHistory` entries by `PostId`, we can count the number of distinct users (`UserId`) who have made an edit for each post. Then we count the rows to get the sum. From below we find that 54293 has several editors.

```
WITH PostEditors AS (
  SELECT PostId, COUNT(DISTINCT(UserId)) AS Editors
  FROM PostHistory
  WHERE PostHistoryTypeId IN (4, 5, 6)
  GROUP BY PostId
  )

SELECT COUNT(*) AS PostMultiEditors
FROM PostEditors
WHERE Editors > 1;
```

Table 19: 1 records

PostMultiEditors
54293

21 Compute the table that contains

- the question,

- the name of the user who posted it,

- when that user joined,

- their location

- the date the question was first posted,

- the accepted answer,

- when the accepted answer was posted

- the name of the user who provided the accepted answer.

The question is found as the `Body` of a Post with `PostTypeId = 1` (question) in the `Posts` table. For the user who posted, it is the `DisplayName` of the, found in the `Users` table by joining on `Id = OwnerUserId` of the post. The user join date, location is simply `CreationDate` and `Location` in the `Users` table. The date the question was first posted is on the other hand `CreationDate` with `PostTypeId = 1` in the `Posts` table. The accepted answer is the Body of a Post with `PostTypeId = 2` (answer) in the `Posts` table which has `Id` equal to the `AcceptedAnswerId` of the question. Following this, the accepted answer date is the `CreationDate` of the accepted answer, found in the `Posts` table. Finaly, the name of the user who provided the accepted answer is the `DisplayName` of the User who posted the accepted answer, found in the `Users` table by joining on `Id = OwnerUserId` of the accepted answer.

```
WITH Answers AS (
  SELECT p.Id, p.Body, p.CreationDate, u.DisplayName
  FROM Posts p
  INNER JOIN Users u ON p.OwnerUserId = u.Id
  WHERE PostTypeId = 2
)

SELECT
  p.Body AS Question,
  u.DisplayName AS UserName,
  u.CreationDate AS UserJoined,
  u.Location AS UserLocation,
  p.CreationDate AS PostDate,
  a.body AS AcceptedAnswer,
  a.CreationDate AS AnswerDate,
  a.DisplayName AS AnswerUser
FROM Posts p
INNER JOIN Users u ON p.OwnerUserId = u.Id
INNER JOIN Answers a ON p.AcceptedAnswerId = a.Id
WHERE p.PostTypeId = 1
```

Table 20: Displaying records 1 - 10

Question	User Name	User Creation Date	User Location	Post Date	Accepted Answer	Answer Date	Answer User
	csgillespie	2010-07-19T[illegible]	Newcastle, United Kingdom	2010-07-19T[illegible]12.510		2010-07-19T19:19:46.160	Harlan
	A Lion	2010-07-19T19:[illegible]		2010-07-19T[illegible]57.157		2010-07-19T[illegible]43:20.423	John L. Taylor
	grokus	2010-07-19T19:[illegible]	United States	2010-07-19T[illegible]28.577		2010-07-19T19:14:43.050	Jay Stevens
	Jay Stevens	2010-07-19T[illegible]	Jacksonville, FL, USA	2010-07-19T[illegible]31.617		2010-07-19T[illegible]36:12.850	John L. Taylor
	EAMann	2010-07-19T[illegible]	Tualatin, OR, United States	2010-07-19T[illegible]59.303		2010-07-19T19:24:18.580	Stephen Turner
	A Lion	2010-07-19T19:[illegible]		2010-07-19T[illegible]47.537		2010-08-19T10:00:00.370	chl
	Christopher D. Long	2010-07-19T19:[illegible]	Versailles, KY	2010-07-19T[illegible]30.810		2010-08-03T[illegible]:50:09.007	Carlos Accioly
	Mehper C. Palavuzlar	2010-07-19T19:[illegible]	Istanbul, Turkiye	2010-07-19T[illegible]04.363		2010-07-19T20:15:54.823	Paul
	Mehper C. Palavuzlar	2010-07-19T19:[illegible]	Istanbul, Turkiye	2010-07-19T[illegible]13.503		2010-07-19T19:29:06.527	Shane
	Oren Hizkiya	2010-07-19T[illegible]	New York, NY	2010-07-19T[illegible]43.860		2010-07-19T19:44:35.037	Neil McGuigan

22 Determine the users that have only posted questions and never answered a question?

The query starts by merging the `Users` and `Posts` tables on `UserId.` The resultant table is then filtered to keep only posts that are either questions or answers, as defined by their `PostTypeId`. For each user, the number of questions they've posted is tallied by counting each instance where the `PostTypeId` equals 1, and similarly, the number of answers is tallied by counting each instance where the `PostTypeId` equals 2. After tallying these values, the result is grouped by `UserId`, ensuring each row in the output corresponds to a unique user. Finally, a `HAVING` clause filters out users who have not posted any questions or have posted at least one answer, thus leaving only those users in the final output who have posted questions and never

posted answers. My quickly browsing the table we find that such users rarely has asked more than one or two questions on the forum, suggesting they are not all to active members.

```
SELECT
  u.DisplayName,
  SUM(CASE WHEN p.PostTypeId = 1 THEN 1 ELSE 0 END) AS Questions,
  SUM(CASE WHEN p.PostTypeId = 2 THEN 1 ELSE 0 END) AS Answers
FROM Users u
INNER JOIN Posts p ON u.Id = p.OwnerUserId
WHERE p.PostTypeId IN (1,2)
GROUP BY u.Id
HAVING Questions > 0 AND Answers = 0
```

Table 21: Displaying records 1 - 10

DisplayName	Questions	Answers
grokus	2	0
A Lion	2	0
EAMann	1	0
Alan H.	13	0
kyle	2	0
Preets	1	0
Martin	2	0
Daniel Vassallo	1	0
Oren Hizkiya	3	0
bshor	1	0

22.1 How many are there?

```
SELECT COUNT(*) AS Posters
FROM (
  SELECT
    u.DisplayName,
    SUM(CASE WHEN p.PostTypeId = 1 THEN 1 ELSE 0 END) AS Questions,
    SUM(CASE WHEN p.PostTypeId = 2 THEN 1 ELSE 0 END) AS Answers
  FROM Users u
  INNER JOIN Posts p ON u.Id = p.OwnerUserId
  WHERE p.PostTypeId IN (1,2)
  GROUP BY u.Id
  HAVING Questions > 0 AND Answers = 0
  )
```

Table 22: 1 records

Posters
76410

23 Compute the table with information for the 75 users with the most accepted answers. This table should include:

- the user's display name,

- creation date,

- location,

- the number of badges they have won, • the names of the badges (as a single string)

- the dates of the earliest and most recent accepted answer (as two fields)

– the (unique) tags for all the questions for which they had the accepted answer (as a single string)

The query starts with a CTE named `AnswerTags`, which compiles a list of unique tags for each user who has posted an answer. This list of tags is aggregated into a single comma-separated string for each user as suggested by Lustig (2010). After that, the query performs a series of joins between the `Posts` and `Users` tables to create a list of users who have had their posts accepted as answers. For each of these users, it calculates several of the asked columns. The results are then joined with the `AnswerTags` to add the string of unique tags associated with each user's accepted answers. Finally, the users are sorted in descending order based on the number of their accepted answers, and the top 75 are selected for the final output.

Glen_b is a strong leader of most accepted answers. Given that he registered in 2010, he is one of the older members of the forum. He's ben receiveng accpeted answers since 2010 and still do today.

```
WITH AnswerTags AS (
  SELECT
    u.Id,
    u.DisplayName,
    GROUP_CONCAT(DISTINCT(tp.tag)) AS UniqueTags
  FROM TagPosts tp
  INNER JOIN Posts p on tp.Id = p.Id
  INNER JOIN Users u ON p.OwnerUserId = u.Id
  GROUP BY u.Id
  )

SELECT
  COUNT(ps.OwnerUserId) AS NumAcceptedAnswers,
  u.DisplayName AS UserName,
  u.CreationDate AS UserCreationDate,
  u.Location,
  MIN(p.CreationDate) AS EarliestAcceptedAnswer,
  MAX(p.CreationDate) AS LatestAcceptedAnswer,
  (SELECT COUNT(UserId) FROM Badges WHERE UserId = ps.OwnerUserId) AS NumBadges,
  (SELECT GROUP_CONCAT(Name, ', ') FROM Badges WHERE UserId = ps.OwnerUserId) AS BadgeNames,
  (SELECT GROUP_CONCAT(Name, ', ') FROM Badges WHERE UserId = ps.OwnerUserId) AS BadgeNames,
  at.UniqueTags
FROM Posts p
INNER JOIN Posts ps ON p.AcceptedAnswerId = ps.Id
INNER JOIN Users u ON ps.OwnerUserId = u.Id
INNER JOIN AnswerTags at ON ps.OwnerUserId = at.Id
GROUP BY ps.OwnerUserId
ORDER BY NumAcceptedAnswers DESC
LIMIT 75;
```

Nu	U	A	N	G	B	N	A	BadgeNames	UniqueTags

Table 23: Displaying records 1 - 10

Nu	U	A	N	G	B	N	A	BadgeNames	UniqueTags
233	5	2010-08-07	2012-07-02	2360	Teacher, Editor, Supporter, Yearling, Commentator, Critic, Student, Scholar, Yearling, Nice Answer, Analytical, Custodian, Custodian, Excavator, Organizer, Enthusiast, Citizen Patrol, Custodian, Tag Editor, Nice Answer, Enlightened, Custodian, Custodian, Custodian, Nice Answer, Enlightened, Informed, Suffrage, Vox Populi, Civic Duty, Quorum, Electorate, Mortarboard, Generalist, Nice Answer, Sportsmanship, Strunk & White, Synonymizer, Nice Answer, regression, Enlightened, Nice Answer, Enlightened, Fanatic, hypothesis-testing, Nice Answer, Good Answer, r, Nice Answer, Revival, Necromancer, Investor, Nice Answer, Nice Answer, Cleanup, Announcer, Booster, Proofreader, Good Answer, Revival, Nice Answer, self-study, Nice Answer, Enlightened, Nice Answer, Pundit, Copy Editor, Revival, Revival, Altruist, Convention, Reviewer, Research Assistant, distributions, Nice Answer, Nice Answer, Enlightened, Nice Answer, Enlightened, histogram, Self-Learner, Talkative, Autobiographer, statistical-significance, Nice Answer, Nice Answer, Enlightened, normal-distribution, Nice Answer, Yearling, data-visualization, t-test, Nice Answer, Nice Answer, Enlightened, Nice Answer, probability, Revival, Nice Answer, Nice Answer, Caucus, Nice Answer, Necromancer, Constituent, Nice Answer, Enlightened, Revival, Revival, data-transformation, Nice Answer, Enlightened, Tumbleweed, Promoter, Revival, Necromancer, Benefactor, Nice Question, Revival, Revival, Good Answer, Nice Answer, Enlightened, p-value, Nice Answer, Enlightened, Reviewer, Revival, Nice Question, Nice Answer, Nice Answer, Enlightened, Good Answer, r, Nice Answer, Enlightened, Nice Answer, Necromancer, Enlightened, Revival, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Enlightened, Enlightened, Nice Question, Nice Answer, Nice Answer, regression,	Teacher, Editor, Supporter, Yearling, Commentator, Critic, Student, Scholar, Yearling, Nice Answer, Analytical, Custodian, Custodian, Excavator, Organizer, Enthusiast, Citizen Patrol, Custodian, Tag Editor, Nice Answer, Enlightened, Custodian, Custodian, Custodian, Nice Answer, Enlightened, Informed, Suffrage, Vox Populi, Civic Duty, Quorum, Electorate, Mortarboard, Generalist, Nice Answer, Sportsmanship, Strunk & White, Synonymizer, Nice Answer, regression, Enlightened, Nice Answer, Enlightened, Fanatic, hypothesis-testing, Nice Answer, Good Answer, r, Nice Answer, Revival, Necromancer, Investor, Nice Answer, Nice Answer, Cleanup, Announcer, Booster, Proofreader, Good Answer, Revival, Nice Answer, self-study, Nice Answer, Enlightened, Nice Answer, Pundit, Copy Editor, Revival, Revival, Altruist, Convention, Reviewer, Research Assistant, distributions, Nice Answer, Nice Answer, Enlightened, Nice Answer, Enlightened, histogram, Self-Learner, Talkative, Autobiographer, statistical-significance, Nice Answer, Nice Answer, Enlightened, normal-distribution, Nice Answer, Yearling, data-visualization, t-test, Nice Answer, Nice Answer, Enlightened, Nice Answer, probability, Revival, Nice Answer, Nice Answer, Caucus, Nice Answer, Necromancer, Constituent, Nice Answer, Enlightened, Revival, Revival, data-transformation, Nice Answer, Enlightened, Tumbleweed, Promoter, Revival, Necromancer, Benefactor, Nice Question, Revival, Revival, Good Answer, Nice Answer, Enlightened, p-value, Nice Answer, Enlightened, Reviewer, Revival, Nice Question, Nice Answer, Nice Answer, Enlightened, Good Answer, r, Nice Answer, Enlightened, Nice Answer, Necromancer, Enlightened, Revival, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Enlightened, Enlightened, Nice Question, Nice Answer, Nice Answer, regression,	partitioning,poisson-process,hypothesis-testing,statistical-significance,chi-squared-test,interpolation,tabl of-fit,binning,model-selection,degrees-of-freedom,rule-of-thumb,algorithms,wilc mann-whitney-test,combinatorics,esti likelihood,mse,method of-moments,efficiency,dis distribution,lognormal distribution,heavy-tailed,correlation,norm distribution,quantiles, variables,observationa study,generalized-linear-model,exponential-family,link-function,logarithmic-series-distribution,time-series,unit-root,stationarity,augm dickey-fuller,regression-coefficients,mathemati statistics,moment-generating-function,saddlepoint-approximation,partial moments,self-study,bernoulli-process,optimization,t distribution,moments, distribution,iid,symme

[illegible] Answer	BadgeNames	UniqueTags
178 whuber 2010-08-13T15:29:47 2010-08-03- 2023- [illegible] Autobiographer, Teacher, Student, Supporter, Editor, Commentator, Organizer, Scholar, Enthusiast, Civic Duty, Tag Editor, Critic, Nice Answer, Mortarboard, Necromancer, Nice Answer, Suffrage, Necromancer, Revival, Revival, Citizen Patrol, Nice Answer, Talkative, Nice Answer, Convention, Quorum, Nice Answer, distributions, Nice Answer, Sportsmanship, Nice Answer, Revival, Revival, Strunk & White, Necromancer, Necromancer, Nice Answer, Enlightened, Nice Answer, Electorate, Nice Answer, Enlightened, Nice Answer, Enlightened, Pundit, Nice Answer, Enlightened, Cleanup, Nice Answer, Nice Answer, Nice Answer, Enlightened, Nice Answer, regression, Nice Answer, Nice Question, Nice Answer, Nice Answer, probability, Nice Answer, Fanatic, hypothesis-testing, Nice Answer, Enlightened, Nice Answer, Nice Answer, Enlightened, Good Answer, Nice Answer, Enlightened, Enlightened, Nice Answer, Enlightened, Synonymizer, Nice Answer, Nice Answer, Nice Answer, Enlightened, Nice Answer, r, Nice Answer, Enlightened, Yearling, Necromancer, Excavator, Nice Answer, Revival, Nice Answer, Nice Answer, normal-distribution, Good Answer, time-series, Investor, Nice Answer, Enlightened, Altruist, Analytical, Nice Answer, Enlightened, correlation, Proofreader, Nice Answer, Disciplined, Revival, Nice Answer, Copy Editor, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Enlightened, Enlightened, Good Answer, Nice Answer, Nice Answer, Nice Answer, Nice Answer, sampling, Nice Answer, Enlightened, Good Answer, Announcer, Nice Answer, Nice Answer, data-visualization, Nice Answer, Taxonomist, Revival, Revival, Guru, Good Answer, Nice Answer, Nice Answer, Guru, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Nice Question, Good Answer, variance, Nice Answer, Nice Answer, Enlightened, Nice Answer, Nice Answer, Good Answer, Good Answer, Nice Answer, Caucus, Constituent, Nice Answer, Nice Answer,	Autobiographer, Teacher, Student, Supporter, Editor, Commentator, Organizer, Scholar, Enthusiast, Civic Duty, Tag Editor, Critic, Nice Answer, Mortarboard, Necromancer, Nice Answer, Suffrage, Necromancer, Revival, Revival, Citizen Patrol, Nice Answer, Talkative, Nice Answer, Convention, Quorum, Nice Answer, distributions, Nice Answer, Sportsmanship, Nice Answer, Revival, Revival, Strunk & White, Necromancer, Necromancer, Nice Answer, Enlightened, Nice Answer, Electorate, Nice Answer, Enlightened, Nice Answer, Enlightened, Pundit, Nice Answer, Enlightened, Cleanup, Nice Answer, Nice Answer, Nice Answer, Enlightened, Nice Answer, regression, Nice Answer, Nice Question, Nice Answer, Nice Answer, probability, Nice Answer, Fanatic, hypothesis-testing, Nice Answer, Enlightened, Nice Answer, Nice Answer, Enlightened, Good Answer, Nice Answer, Enlightened, Enlightened, Nice Answer, Enlightened, Synonymizer, Nice Answer, Nice Answer, Nice Answer, Enlightened, Nice Answer, r, Nice Answer, Enlightened, Yearling, Necromancer, Excavator, Nice Answer, Revival, Nice Answer, Nice Answer, normal-distribution, Good Answer, time-series, Investor, Nice Answer, Enlightened, Altruist, Analytical, Nice Answer, Enlightened, correlation, Proofreader, Nice Answer, Disciplined, Revival, Nice Answer, Copy Editor, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Enlightened, Enlightened, Good Answer, Nice Answer, Nice Answer, Nice Answer, Nice Answer, sampling, Nice Answer, Enlightened, Good Answer, Announcer, Nice Answer, Nice Answer, data-visualization, Nice Answer, Taxonomist, Revival, Revival, Guru, Good Answer, Nice Answer, Nice Answer, Guru, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Nice Question, Good Answer, variance, Nice Answer, Nice Answer, Enlightened, Nice Answer, Nice Answer, Good Answer, Good Answer, Nice Answer, Caucus, Constituent, Nice Answer, Nice Answer,	correlation,estimation matrix,unbiased-estimator,modeling,m analysis,spatial,decisio theory,paradox,probab processes,puzzle,hypot testing,statistical-significance,p-value,confidence-interval,survey,small-sample,stratification,h distribution,distributi transformation,entrop regression,likelihood,s parameter,location-parameter

Number	UserId	CreationDate	Text	BadgeNames	UniqueTags
11	9	2010-08-12T04:59:29Z	Teacher, Editor, Supporter, Critic, Commentator, Nice Answer, Custodian, Explainer, Organizer, Enthusiast, Autobiographer, Custodian, Fanatic, Custodian, Custodian, Nice Answer, Mortarboard, Tenacious, Custodian, Yearling, probability, Nice Answer, Enlightened, Citizen Patrol, Nice Answer, Enlightened, Student, Scholar, Nice Answer, Enlightened, self-study, machine-learning, Nice Answer, Good Answer, Nice Answer, Enlightened, Nice Answer, Enlightened, mathematical-statistics, Nice Answer, Enlightened, Generalist, Civic Duty, normal-distribution, probability, Nice Answer, Nice Answer, Enlightened, distributions, Enlightened, regression, expected-value, neural-networks, conditional-probability, random-variable, Yearling, Nice Answer, Enlightened, variance, Good Answer, Nice Answer, bayesian, Nice Answer, Enlightened, Electorate, Nice Answer, Nice Answer, Enlightened, Good Answer, Nice Answer, Enlightened, Guru, independence, Nice Answer, Nice Answer, Nice Answer, Enlightened, Nice Answer, Enlightened, Strunk & White, machine-learning, classification, Nice Answer, covariance, Enlightened, Nice Answer, Caucus, Constituent, Nice Answer, Nice Answer, Enlightened, Nice Answer, Refiner, probability, Nice Answer, Enlightened, time-series, Nice Answer, Yearling, Nice Answer, Good Answer, Nice Answer, Enlightened, Good Answer, Nice Answer, cross-validation, Nice Answer, Enlightened, Nice Answer, Nice Answer, conditional-expectation, Nice Answer, Enlightened, Enlightened, Nice Answer, Nice Answer, optimization, Nice Answer, Yearling, Enlightened, Nice Answer, scikit-learn, Nice Answer, Pundit, Enlightened, Nice Answer, machine-learning, Enlightened, neural-networks	Teacher, Editor, Supporter, Critic, Commentator, Nice Answer, Custodian, Explainer, Organizer, Enthusiast, Autobiographer, Custodian, Fanatic, Custodian, Custodian, Nice Answer, Mortarboard, Tenacious, Custodian, Yearling, probability, Nice Answer, Enlightened, Citizen Patrol, Nice Answer, Enlightened, Student, Scholar, Nice Answer, Enlightened, self-study, machine-learning, Nice Answer, Good Answer, Nice Answer, Enlightened, Nice Answer, Enlightened, mathematical-statistics, Nice Answer, Enlightened, Generalist, Civic Duty, normal-distribution, probability, Nice Answer, Nice Answer, Enlightened, distributions, Enlightened, regression, expected-value, neural-networks, conditional-probability, random-variable, Yearling, Nice Answer, Enlightened, variance, Good Answer, Nice Answer, bayesian, Nice Answer, Enlightened, Electorate, Nice Answer, Nice Answer, Enlightened, Good Answer, Nice Answer, Enlightened, Guru, independence, Nice Answer, Nice Answer, Nice Answer, Enlightened, Nice Answer, Enlightened, Strunk & White, machine-learning, classification, Nice Answer, covariance, Enlightened, Nice Answer, Caucus, Constituent, Nice Answer, Nice Answer, Enlightened, Nice Answer, Refiner, probability, Nice Answer, Enlightened, time-series, Nice Answer, Yearling, Nice Answer, Good Answer, Nice Answer, Enlightened, Good Answer, Nice Answer, cross-validation, Nice Answer, Enlightened, Nice Answer, Nice Answer, conditional-expectation, Nice Answer, Enlightened, Enlightened, Nice Answer, Nice Answer, optimization, Nice Answer, Yearling, Enlightened, Nice Answer, scikit-learn, Nice Answer, Pundit, Enlightened, Nice Answer, machine-learning, Enlightened, neural-networks	probability,uniform-distribution

Nu	U	A	N	G	B	N	A	BadgeNames	UniqueTags
100	10	13	2	2	2	2	7	Editor, Autobiographer, Supporter, Teacher, Custodian, Student, Scholar, Excavator, Commentator, Citizen Patrol, Critic, Explainer, Revival, Custodian, Nice Answer, Custodian, Nice Answer, Tumbleweed, Quorum, Informed, Yearling, Nice Answer, Nice Answer, Enlightened, Curious, Enthusiast, Custodian, Civic Duty, Nice Answer, Revival, Nice Answer, Enlightened, Revival, Custodian, Organizer, Nice Answer, Suffrage, Nice Answer, Nice Question, Strunk & White, Fanatic, Nice Answer, Enlightened, Caucus, Proofreader, Custodian, Cleanup, Constituent, Convention, Deputy, Nice Answer, Enlightened, Nice Answer, Necromancer, Electorate, Popular Question, Nice Answer, Tag Editor, Reviewer, Pundit, Taxonomist, Nice Answer, Enlightened, Nice Answer, Nice Answer, Good Answer, Nice Answer, Mortarboard, Enlightened, Nice Answer, Good Answer, Revival, r, Generalist, Revival, Nice Answer, Yearling, Enlightened, Good Answer, bayesian, Revival, Necromancer, Necromancer, Nice Answer, Enlightened, Nice Answer, Self-Learner, Nice Answer, Enlightened, Revival, Necromancer, Nice Answer, Enlightened, Nice Question, Notable Question, Good Question, Promoter, Nice Answer, Benefactor, Necromancer, Nice Answer, Nice Answer, Enlightened, Good Answer, Sportsmanship, Nice Question, Favorite Question, Popular Question, Nice Answer, Popular Question, Good Question, regression, Nice Answer, distributions, Enlightened, Enlightened, normal-distribution, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Enlightened, Nice Question, Popular Question, Good Question, Notable Question, Favorite Question, Great Question, Famous Question, Guru, Nice Answer, Enlightened, Nice Answer, probability, Nice Answer, Refiner, Nice Question, Nice Answer, Inquisitive, Nice Question, Popular Question, Good Question, Notable Question, Favorite Question, Nice Question, Nice Answer, Enlightened, Nice Answer, Necromancer, Revival, terminology, Nice Answer,	Editor, Autobiographer, Supporter, Teacher, Custodian, Student, Scholar, Excavator, Commentator, Citizen Patrol, Critic, Explainer, Revival, Custodian, Nice Answer, Custodian, Nice Answer, Tumbleweed, Quorum, Informed, Yearling, Nice Answer, Nice Answer, Enlightened, Curious, Enthusiast, Custodian, Civic Duty, Nice Answer, Revival, Nice Answer, Enlightened, Revival, Custodian, Organizer, Nice Answer, Suffrage, Nice Answer, Nice Question, Strunk & White, Fanatic, Nice Answer, Enlightened, Caucus, Proofreader, Custodian, Cleanup, Constituent, Convention, Deputy, Nice Answer, Enlightened, Nice Answer, Necromancer, Electorate, Popular Question, Nice Answer, Tag Editor, Reviewer, Pundit, Taxonomist, Nice Answer, Enlightened, Nice Answer, Nice Answer, Good Answer, Nice Answer, Mortarboard, Enlightened, Nice Answer, Good Answer, Revival, r, Generalist, Revival, Nice Answer, Yearling, Enlightened, Good Answer, bayesian, Revival, Necromancer, Necromancer, Nice Answer, Enlightened, Nice Answer, Self-Learner, Nice Answer, Enlightened, Revival, Necromancer, Nice Answer, Enlightened, Nice Question, Notable Question, Good Question, Promoter, Nice Answer, Benefactor, Necromancer, Nice Answer, Nice Answer, Enlightened, Good Answer, Sportsmanship, Nice Question, Favorite Question, Popular Question, Nice Answer, Popular Question, Good Question, regression, Nice Answer, distributions, Enlightened, Enlightened, normal-distribution, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Enlightened, Nice Question, Popular Question, Good Question, Notable Question, Favorite Question, Great Question, Famous Question, Guru, Nice Answer, Enlightened, Nice Answer, probability, Nice Answer, Refiner, Nice Question, Nice Answer, Inquisitive, Nice Question, Popular Question, Good Question, Notable Question, Favorite Question, Nice Question, Nice Answer, Enlightened, Nice Answer, Necromancer, Revival, terminology, Nice Answer,	r,mixed-model,lme4-nlme,model-selection,multilevel-analysis,intraclass-correlation,modeling,r statistics,definition,da visualization,time-series,forecasting,smal sample,references,regr model,robust,simulati carlo,metropolis-hastings,genetic-algorithms,evolutionar algorithms,error,predi interval,bayesian,gene linear-model,random-generation,kernel-smoothing,binomial-distribution,proportio distribution,hypergeor distribution,hypothesi testing,p-value,frequentist,distri theory,confidence-interval,precision,mac learning,forecastabilit chain-montecarlo,cumulative distribution-function,moment-generating-function,philosophical paradox,probability,di distribution,uniform-distribution,triangular distribution,truncatio point,algorithms,mult distribution,history,be binomial-distribution,loss-functions,decision-theory,multivariate-analysis,kolmogorov-smirnov-test,empirical-cumulative-distr-fn,terminology,iid,nega binomial-distribution,zero-inflation,classification models,unbalanced-classes,faq,order-statistics,extreme-value,mixture-distribution,density-estimation,nonparame

NuUsersAnNapCrEditionNDaRepNaAnswer	BadgeNames	UniqueTags
985Step2018wi2020-238 Ko-09- 09-03- lassa18T1023505:3846:20493 Student, Teacher, Supporter, Autobiographer, Editor, Scholar, Nice Answer, Commentator, Organizer, Enthusiast, Nice Answer, Suffrage, Nice Answer, Yearling, Popular Question, Nice Answer, Good Answer, Nice Answer, Analytical, Critic, Yearling, Nice Question, Nice Answer, Custodian, Disciplined, Notable Question, Necromancer, Citizen Patrol, Custodian, Civic Duty, Tag Editor, Custodian, Custodian, Custodian, Nice Answer, Fanatic, Revival, Promoter, Custodian, Good Answer, Yearling, Nice Answer, Pundit, Nice Answer, Nice Answer, Nice Answer, Yearling, time-series, Explainer, Informed, Nice Answer, Nice Answer, forecasting, Quorum, Good Answer, Caucus, Constituent, r, Excavator, regression, Nice Answer, Nice Answer, Enlightened, Nice Answer, Nice Answer, Nice Answer, Good Answer, Yearling, Nice Answer, Generalist, Strunk & White, Nice Answer, Enlightened, Nice Answer, Revival, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Enlightened, Nice Answer, Convention, Nice Answer, Necromancer, Cleanup, Nice Answer, Enlightened, Good Answer, Electorate, Custodian, Talkative, Investor, Refiner, Altruist, Nice Answer, Sportsmanship, arima, Outspoken, Nice Answer, Enlightened, Nice Answer, Mortarboard, Good Answer, forecasting, Great Answer, Nice Answer, Good Answer, Guru, Taxonomist, Guru, Enlightened, Yearling, Nice Answer, machine-learning, Nice Answer, Enlightened, Nice Answer, Guru, Nice Answer, Enlightened, Nice Answer, Enlightened, time-series, Nice Answer, Enlightened, Good Answer, Nice Answer, Enlightened, Necromancer, Synonymizer, Necromancer, Nice Answer, Enlightened, Nice Answer, Enlightened, Announcer, Announcer, Booster, data-visualization, Nice Answer, Nice Answer, Good Answer, Announcer, Nice Answer, Caucus, Good Answer, Booster, Nice Answer, Constituent, Enlightened, Self-Learner, Curious, Yearling, Reviewer, Nice Answer, Enlightened, distributions, Nice Answer, Nice Answer, Nice Answer, Nice Answer	Student, Teacher, Supporter, Autobiographer, Editor, Scholar, Nice Answer, Commentator, Organizer, Enthusiast, Nice Answer, Suffrage, Nice Answer, Yearling, Popular Question, Nice Answer, Good Answer, Nice Answer, Analytical, Critic, Yearling, Nice Question, Nice Answer, Custodian, Disciplined, Notable Question, Necromancer, Citizen Patrol, Custodian, Civic Duty, Tag Editor, Custodian, Custodian, Custodian, Nice Answer, Fanatic, Revival, Promoter, Custodian, Good Answer, Yearling, Nice Answer, Pundit, Nice Answer, Nice Answer, Nice Answer, Yearling, time-series, Explainer, Informed, Nice Answer, Nice Answer, forecasting, Quorum, Good Answer, Caucus, Constituent, r, Excavator, regression, Nice Answer, Nice Answer, Enlightened, Nice Answer, Nice Answer, Nice Answer, Good Answer, Yearling, Nice Answer, Generalist, Strunk & White, Nice Answer, Enlightened, Nice Answer, Revival, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Enlightened, Nice Answer, Convention, Nice Answer, Necromancer, Cleanup, Nice Answer, Enlightened, Good Answer, Electorate, Custodian, Talkative, Investor, Refiner, Altruist, Nice Answer, Sportsmanship, arima, Outspoken, Nice Answer, Enlightened, Nice Answer, Mortarboard, Good Answer, forecasting, Great Answer, Nice Answer, Good Answer, Guru, Taxonomist, Guru, Enlightened, Yearling, Nice Answer, machine-learning, Nice Answer, Enlightened, Nice Answer, Guru, Nice Answer, Enlightened, Nice Answer, Enlightened, time-series, Nice Answer, Enlightened, Good Answer, Nice Answer, Enlightened, Necromancer, Synonymizer, Necromancer, Nice Answer, Enlightened, Nice Answer, Enlightened, Announcer, Announcer, Booster, data-visualization, Nice Answer, Nice Answer, Good Answer, Announcer, Nice Answer, Caucus, Good Answer, Booster, Nice Answer, Constituent, Enlightened, Self-Learner, Curious, Yearling, Reviewer, Nice Answer, Enlightened, distributions, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Nice Answer	correlation,multiple-comparisons,statistica significance,genetics,cl measures,missing-data,regularization,ra distribution,entropy,n binomial-distribution,time-series,data-visualization,barplot,c data,ranking,accuracy classes,oversampling,f distribution,random-generation,sum,discre data,distribution-identification,underdis maxwell-poisson-distribution,forecastin distribution,mse,mae, distribution,distributi of-fit,median,r,copula,m distribution,reference distribution,probabilit rules,highest-density-region,random-forest,arima,model-selection,box-jenkins,information-criteria,sensitivity-specificity,f1,confusion matrix

Nu	U	A	N	G	B	N	A	BadgeNames	UniqueTags
984	Xi'an	201	Pa	201	273	Teacher, Editor, Supporter, Critic, Commentator, Analytical, Autobiographer, Organizer, Student, Announcer, Announcer, Scholar, Nice Answer, Enlightened, Enthusiast, Disciplined, Tag Editor, Citizen Patrol, Announcer, Civic Duty, Announcer, Announcer, Announcer, Promoter, Self-Learner, Benefactor, Announcer, Revival, Excavator, Nice Answer, Nice Question, Nice Answer, Nice Question, Revival, Nice Answer, Announcer, Announcer, Yearling, Nice Answer, Enlightened, bayesian, Revival, Mortarboard, Revival, Revival, Revival, Revival, Nice Question, Yearling, Nice Answer, Good Answer, Taxonomist, Nice Answer, Explainer, Yearling, Announcer, Good Answer, Announcer, Custodian, Custodian, Nice Question, Custodian, Custodian, Revival, Announcer, Announcer, Announcer, Custodian, Custodian, Revival, Revival, Strunk & White, Necromancer, Necromancer, Necromancer, Nice Answer, Revival, Announcer, Informed, Revival, Revival, Revival, Nice Answer, Nice Answer, Revival, Nice Answer, Reviewer, Announcer, distributions, Announcer, Nice Answer, Enlightened, Electorate, Announcer, Fanatic, Nice Answer, Nice Answer, Announcer, Necromancer, Announcer, Caucus, Constituent, Pundit, Reviewer, Nice Answer, Enlightened, Nice Answer, mathematical-statistics, self-study, Announcer, Popular Question, Necromancer, Refiner, Proofreader, Enlightened, Announcer, Announcer, Nice Answer, Steward, sampling, Nice Answer, Curious, Nice Answer, Enlightened, Revival, Revival, Reviewer, Nice Answer, Reviewer, Announcer, Nice Answer, simulation, probability, Revival, Announcer, monte-carlo, Yearling, Guru, Necromancer, Revival, Nice Answer, Announcer, Revival, Announcer, Nice Answer, bayesian, Nice Answer, Enlightened, Nice Answer, Cleanup, Nice Answer, r, Sportsmanship, Investor, Announcer, Nice Answer, normal-distribution, Announcer, Good Answer, Nice Answer, Enlightened, Announcer, gibbs, Nice Answer, Announcer, Nice Answer, Nice	Teacher, Editor, Supporter, Critic, Commentator, Analytical, Autobiographer, Organizer, Student, Announcer, Announcer, Scholar, Nice Answer, Enlightened, Enthusiast, Disciplined, Tag Editor, Citizen Patrol, Announcer, Civic Duty, Announcer, Announcer, Announcer, Promoter, Self-Learner, Benefactor, Announcer, Revival, Excavator, Nice Answer, Nice Question, Nice Answer, Nice Question, Revival, Nice Answer, Announcer, Announcer, Yearling, Nice Answer, Enlightened, bayesian, Revival, Mortarboard, Revival, Revival, Revival, Revival, Nice Question, Yearling, Nice Answer, Good Answer, Taxonomist, Nice Answer, Explainer, Yearling, Announcer, Good Answer, Announcer, Custodian, Custodian, Nice Question, Custodian, Custodian, Revival, Announcer, Announcer, Announcer, Custodian, Custodian, Revival, Revival, Strunk & White, Necromancer, Necromancer, Necromancer, Nice Answer, Revival, Announcer, Informed, Revival, Revival, Revival, Nice Answer, Nice Answer, Revival, Nice Answer, Reviewer, Announcer, distributions, Announcer, Nice Answer, Enlightened, Electorate, Announcer, Fanatic, Nice Answer, Nice Answer, Announcer, Necromancer, Announcer, Caucus, Constituent, Pundit, Reviewer, Nice Answer, Enlightened, Nice Answer, mathematical-statistics, self-study, Announcer, Popular Question, Necromancer, Refiner, Proofreader, Enlightened, Announcer, Announcer, Nice Answer, Steward, sampling, Nice Answer, Curious, Nice Answer, Enlightened, Revival, Revival, Reviewer, Nice Answer, Reviewer, Announcer, Nice Answer, simulation, probability, Revival, Announcer, monte-carlo, Yearling, Guru, Necromancer, Revival, Nice Answer, Announcer, Revival, Announcer, Nice Answer, bayesian, Nice Answer, Enlightened, Nice Answer, Cleanup, Nice Answer, r, Sportsmanship, Investor, Announcer, Nice Answer, normal-distribution, Announcer, Good Answer, Nice Answer, Enlightened, Announcer, gibbs, Nice Answer, Announcer, Nice Answer, Nice	time-series,polynomial,opti data,expectation-maximization,weibull-distribution,gumbel-distribution,distributi analysis,covariance,de function,unbiased-estimator,distance-functions,functional-data-analysis,hellinger,med family,sufficient-statistics,chemistry,cu distribution-function,mad,simulati chain-montecarlo,copula,bay distribution,sequentia analysis,bayes-factors,monte-carlo,geometric-mean,scalability,finite- mixture-model,matrix,random- generation,wishart-distribution,maximum likelihood,conditional-probability,random-variable,t-distribution,expected-value,conditional-expectation,sampling, statistics,metropolis-hastings,algorithms,gi annealing,chi-squared-distribution,transform gaussian-distribution

Nu	U	A	N	G	E	B	N	A	BadgeNames	UniqueTags
872	Ric Hay	20 08- 08T10	20 02-03- 23T22	24	20	26	3	Student, Scholar, Editor, Commentator, Supporter, Excavator, Teacher, Revival, Autobiographer, Curious, Informed, Tumbleweed, Enthusiast, Organizer, Nice Answer, Enlightened, Fanatic, Quorum, Explainer, Caucus, Constituent, Tenacious, Custodian, Revival, Strunk & White, Critic, Unsung Hero, Nice Answer, Yearling, Civic Duty, Nice Question, time-series, Popular Question, Cleanup, Popular Question, Inquisitive, Nice Answer, Citizen Patrol, Refiner, Proofreader, Pundit, Nice Answer, Copy Editor, Necromancer, Popular Question, r, Revival, Nice Answer, Tag Editor, Revival, regression, Notable Question, Revival, Reviewer, Nice Question, arima, Notable Question, Custodian, Custodian, Custodian, Custodian, Custodian, Nice Answer, Revival, Revival, Nice Question, Popular Question, Revival, Convention, Electorate, Suffrage, Vox Populi, Yearling, garch, Popular Question, forecasting, Nice Answer, Nice Question, Synonymizer, Nice Question, Nice Question, Nice Answer, Good Answer, Mortarboard, Popular Question, Revival, Necromancer, Popular Question, Talkative, Revival, Self-Learner, Revival, Revival, Notable Question, Revival, Revival, time-series, Revival, Nice Question, Nice Answer, Enlightened, Good Answer, Necromancer, Popular Question, Popular Question, Revival, Nice Answer, Nice Question, Popular Question, Revival, Revival, Caucus, Yearling, Necromancer, Revival, Constituent, Generalist, Notable Question, Revival, Popular Question, Popular Question, Popular Question, Taxonomist, Revival, Popular Question, Nice Question, Notable Question, Famous Question, Nice Question, cointegration, Popular Question, Revival, Notable Question, Nice Question, Necromancer, Popular Question, Famous Question, Revival, Good Question, Popular Question, Popular Question, Nice Answer, Revival, Notable Question, Nice Answer, Nice Question, Popular Question, Revival, Popular Question, Yearling, Archaeologist, Popular Question, Nice Question, Nice Answer, Enlightened, Nice Answer	Student, Scholar, Editor, Commentator, Supporter, Excavator, Teacher, Revival, Autobiographer, Curious, Informed, Tumbleweed, Enthusiast, Organizer, Nice Answer, Enlightened, Fanatic, Quorum, Explainer, Caucus, Constituent, Tenacious, Custodian, Revival, Strunk & White, Critic, Unsung Hero, Nice Answer, Yearling, Civic Duty, Nice Question, time-series, Popular Question, Cleanup, Popular Question, Inquisitive, Nice Answer, Citizen Patrol, Refiner, Proofreader, Pundit, Nice Answer, Copy Editor, Necromancer, Popular Question, r, Revival, Nice Answer, Tag Editor, Revival, regression, Notable Question, Revival, Reviewer, Nice Question, arima, Notable Question, Custodian, Custodian, Custodian, Custodian, Custodian, Nice Answer, Revival, Revival, Nice Question, Popular Question, Revival, Convention, Electorate, Suffrage, Vox Populi, Yearling, garch, Popular Question, forecasting, Nice Answer, Nice Question, Synonymizer, Nice Question, Nice Question, Nice Answer, Good Answer, Mortarboard, Popular Question, Revival, Necromancer, Popular Question, Talkative, Revival, Self-Learner, Revival, Revival, Notable Question, Revival, Revival, time-series, Revival, Nice Question, Nice Answer, Enlightened, Good Answer, Necromancer, Popular Question, Popular Question, Revival, Nice Answer, Nice Question, Popular Question, Revival, Revival, Caucus, Yearling, Necromancer, Revival, Constituent, Generalist, Notable Question, Revival, Popular Question, Popular Question, Popular Question, Taxonomist, Revival, Popular Question, Nice Question, Notable Question, Famous Question, Nice Question, cointegration, Popular Question, Revival, Notable Question, Nice Question, Necromancer, Popular Question, Famous Question, Revival, Good Question, Popular Question, Popular Question, Nice Answer, Revival, Notable Question, Nice Answer, Nice Question, Popular Question, Revival, Popular Question, Yearling, Archaeologist, Popular Question, Nice Question, Nice Answer, Enlightened, Nice Answer	aic,arma,garch,volatili forecasting,forecasting series,modeling,factor- analysis,cointegration, interval,bootstrap,qua value,regression,cross- validation,r- squared,intuition,r,mo selection,arima,season error- correction- model,accuracy,mase,v autoregression,statisti significance,sample- size,effect- size,machine- learning,kernel- smoothing,large- data,covariance- matrix,generalized- least- squares,efficiency,mse, testing,references,norm assumption,small- sample,arfima,least- squares,assumptions,a root,random- walk,autocorrelation,la regression,multivariate analysis,arch,estimato variable,terminology,d net,model,definition,b selection,algorithms,ti complexity,regression- strategies,explanatory models,maximum- likelihood,unbiased- estimator,quasi- maximum- likelihood,consistency, value,panel- data,fixed- effects- model,macroeconomic algorithms,degrees- of- freedom,pca,non- independent,iid,oracle power,type- i- and- ii- errors,kalman- filter,state- space- models,distributions,q regression,inference,m statistics,convergence,

Number	UserId	CreationDate	DeletionDate	Text	BadgeNames	UniqueTags
826	3[illegible]	2017-[illegible]	[illegible]	Teacher, Autobiographer, Editor, Revival, Mortarboard, Supporter, Revival, Revival, Commentator, Critic, Revival, Revival, Citizen Patrol, Custodian, Informed, Revival, Student, Nice Question, Revival, Revival, Necromancer, Scholar, Quorum, Enthusiast, Revival, Revival, Revival, Explainer, Revival, Revival, Necromancer, Revival, Revival, Revival, Revival, Custodian, Revival, Nice Answer, Revival, Revival, Revival, Revival, Revival, Revival, Nice Answer, Nice Answer, Fanatic, Revival, Nice Answer, Enlightened, Revival, Good Answer, Revival, Revival, Revival, Nice Answer, probability, Civic Duty, Nice Answer, Revival, Nice Answer, Good Answer, Revival, Enlightened, Revival, Nice Answer, Nice Answer, Enlightened, Necromancer, Necromancer, Necromancer, Necromancer, Revival, Nice Answer, Enlightened, bayesian, Revival, Nice Answer, Enlightened, Suffrage, Generalist, Nice Answer, Revival, Nice Answer, Good Answer, Curious, Revival, Promoter, Guru, Enlightened, Custodian, Benefactor, Yearling, Excavator, Organizer, Custodian, Necromancer, Revival, Cleanup, regression, Vox Populi, Revival, Revival, Revival, Revival, Revival, Disciplined, Nice Answer, Enlightened, Electorate, self-study, Revival, Nice Answer, distributions, Revival, Revival, Revival, Revival, Revival, Nice Answer, Strunk & White, Necromancer, Nice Answer, Self-Learner, mathematical-statistics, Revival, normal-distribution, Revival, Revival, Revival, Revival, Revival, Revival, Revival, Revival, maximum-likelihood, Revival, Nice Answer, hypothesis-testing, Nice Answer, Enlightened, Nice Answer, Good Answer, Nice Answer, random-variable, Necromancer, Nice Question, Nice Answer, Revival, Revival, Revival, Revival, Nice Answer, Enlightened, Good Answer, Guru, Necromancer, Revival, Nice Answer, Enlightened, Sportsmanship, Revival, conditional-probability, Revival, Revival, Necromancer, Revival, r, Nice Answer, Necromancer, Revival, Necromancer, Nice Answer, time-series, Nice Answer, Enlightened,	Teacher, Autobiographer, Editor, Revival, Mortarboard, Supporter, Revival, Revival, Commentator, Critic, Revival, Revival, Citizen Patrol, Custodian, Informed, Revival, Student, Nice Question, Revival, Revival, Necromancer, Scholar, Quorum, Enthusiast, Revival, Revival, Revival, Explainer, Revival, Revival, Necromancer, Revival, Revival, Revival, Revival, Custodian, Revival, Nice Answer, Revival, Revival, Revival, Revival, Revival, Revival, Nice Answer, Nice Answer, Fanatic, Revival, Nice Answer, Enlightened, Revival, Good Answer, Revival, Revival, Revival, Nice Answer, probability, Civic Duty, Nice Answer, Revival, Nice Answer, Good Answer, Revival, Enlightened, Revival, Nice Answer, Nice Answer, Enlightened, Necromancer, Necromancer, Necromancer, Necromancer, Revival, Nice Answer, Enlightened, bayesian, Revival, Nice Answer, Enlightened, Suffrage, Generalist, Nice Answer, Revival, Nice Answer, Good Answer, Curious, Revival, Promoter, Guru, Enlightened, Custodian, Benefactor, Yearling, Excavator, Organizer, Custodian, Necromancer, Revival, Cleanup, regression, Vox Populi, Revival, Revival, Revival, Revival, Revival, Disciplined, Nice Answer, Enlightened, Electorate, self-study, Revival, Nice Answer, distributions, Revival, Revival, Revival, Revival, Revival, Nice Answer, Strunk & White, Necromancer, Nice Answer, Self-Learner, mathematical-statistics, Revival, normal-distribution, Revival, Revival, Revival, Revival, Revival, Revival, Revival, Revival, maximum-likelihood, Revival, Nice Answer, hypothesis-testing, Nice Answer, Enlightened, Nice Answer, Good Answer, Nice Answer, random-variable, Necromancer, Nice Question, Nice Answer, Revival, Revival, Revival, Revival, Nice Answer, Enlightened, Good Answer, Guru, Necromancer, Revival, Nice Answer, Enlightened, Sportsmanship, Revival, conditional-probability, Revival, Revival, Necromancer, Revival, r, Nice Answer, Necromancer, Revival, Necromancer, Nice Answer, time-series, Nice Answer, Enlightened,	probability,distributio transformation,matrix processing,predictor,ra allocation,blocking,sa statistics,t-distribution,multinom distribution,regression linear-model,glmm,nested-data,faq,r,computatio statistics,underflow,lo squared-test,poisson-regression,overdispersi likelihood,forecasting, carlo,rejection-sampling,hypothesis-testing,causality,philos function,random-generation,concavity,s parameter,moments,c series,discrete-distributions,extreme-value,estimation,stand deviation,count-data,poisson-distribution,underdisp distribution,expected-value,binomial-distribution,optimizat size,stratification,rule-of-thumb,confidence-interval,weighted-mean,random-walk,bayesian,interpre population,intuition,s theory,maximum-likelihood,r-squared,aic,bic
	10	08-Aug	12-02-	[illegible]		

[illegible]		BadgeNames	UniqueTags
814 [illegible] Bay Area	Autobiographer, Teacher, Editor, Revival, Revival, Supporter, Commentator, Caucus, Yearling, Critic, Explainer, Citizen Patrol, Yearling, Nice Answer, Nice Answer, Informed, Revival, Enthusiast, Custodian, Yearling, Organizer, Nice Answer, Custodian, Enlightened, Fanatic, Necromancer, Civic Duty, Revival, Nice Answer, Nice Answer, Enlightened, hypothesis-testing, Nice Answer, Nice Answer, Nice Answer, Yearling, normal-distribution, Custodian, distributions, Generalist, probability, r, t-test, Sportsmanship, confidence-interval, statistical-significance, mathematical-statistics, Student, Self-Learner, Nice Answer, Mortarboard, Nice Answer, Good Answer, self-study, Nice Answer, Nice Answer, Electorate, Nice Answer, Enlightened, Pundit, Nice Answer, Enlightened, Good Answer, Enlightened, Nice Answer, Nice Answer, Nice Answer, estimation, mean, Nice Answer, Strunk & White, Yearling, Nice Answer, Enlightened, Nice Answer, anova, bayesian, Caucus, Constituent, variance, Nice Answer, inference, Enlightened, hypothesis-testing, binomial-distribution, Refiner, Nice Answer, p-value, chi-squared-test, Nice Answer, Nice Answer, Good Answer, probability, Nice Answer, Nice Answer, Yearling, Nice Answer, Nice Answer, Nice Answer, sample-size, standard-deviation, Nice Answer, Cleanup, nonparametric, Nice Answer, Enlightened, descriptive-statistics, Nice Answer, Excavator, sampling, poisson-distribution, Disciplined, wilcoxon-mann-whitney-test, Yearling, Good Answer, Nice Question, Nice Answer, Enlightened, Nice Answer, r, proportion, Nice Answer	Autobiographer, Teacher, Editor, Revival, Revival, Supporter, Commentator, Caucus, Yearling, Critic, Explainer, Citizen Patrol, Yearling, Nice Answer, Nice Answer, Informed, Revival, Enthusiast, Custodian, Yearling, Organizer, Nice Answer, Custodian, Enlightened, Fanatic, Necromancer, Civic Duty, Revival, Nice Answer, Nice Answer, Enlightened, hypothesis-testing, Nice Answer, Nice Answer, Nice Answer, Yearling, normal-distribution, Custodian, distributions, Generalist, probability, r, t-test, Sportsmanship, confidence-interval, statistical-significance, mathematical-statistics, Student, Self-Learner, Nice Answer, Mortarboard, Nice Answer, Good Answer, self-study, Nice Answer, Nice Answer, Electorate, Nice Answer, Enlightened, Pundit, Nice Answer, Enlightened, Good Answer, Enlightened, Nice Answer, Nice Answer, Nice Answer, estimation, mean, Nice Answer, Strunk & White, Yearling, Nice Answer, Enlightened, Nice Answer, anova, bayesian, Caucus, Constituent, variance, Nice Answer, inference, Enlightened, hypothesis-testing, binomial-distribution, Refiner, Nice Answer, p-value, chi-squared-test, Nice Answer, Nice Answer, Good Answer, probability, Nice Answer, Nice Answer, Yearling, Nice Answer, Nice Answer, Nice Answer, sample-size, standard-deviation, Nice Answer, Cleanup, nonparametric, Nice Answer, Enlightened, descriptive-statistics, Nice Answer, Excavator, sampling, poisson-distribution, Disciplined, wilcoxon-mann-whitney-test, Yearling, Good Answer, Nice Question, Nice Answer, Enlightened, Nice Answer, r, proportion, Nice Answer	bayesian,estimation,m process

Nu	[illegible]	BadgeNames	UniqueTags
718	Pe 20 N 20 20 18 T Flo 08 Yo 12 07 03 N 12 T 08 1	Teacher, Supporter, Organizer, Student, Scholar, Quorum, Editor, Commentator, Nice Answer, Autobiographer, Necromancer, Necromancer, Yearling, Nice Answer, Critic, Enthusiast, Nice Answer, Nice Answer, Necromancer, Good Answer, Nice Question, regression, Nice Answer, Popular Question, Civic Duty, Nice Answer, Nice Answer, Necromancer, Sportsmanship, Nice Answer, Nice Answer, Necromancer, Citizen Patrol, Necromancer, Nice Answer, Cleanup, Yearling, r, Nice Answer, Strunk & White, Custodian, logistic, Custodian, Revival, Custodian, Custodian, Analytical, Suffrage, Vox Populi, Tag Editor, Excavator, Custodian, Nice Question, correlation, Pundit, Nice Answer, Nice Answer, Custodian, Convention, Reviewer, Necromancer, Generalist, Fanatic, Good Question, Electorate, Proofreader, Custodian, Nice Answer, Reviewer, Reviewer, regression, Nice Answer, Mortarboard, Nice Answer, Copy Editor, multiple-regression, Good Answer, Nice Answer, Necromancer, statistical-significance, hypothesis-testing, Disciplined, Popular Question, Nice Question, Yearling, Nice Answer, Enlightened, Nice Answer, Popular Question, Nice Answer, Enlightened, Caucus, interaction, Constituent, anova, Popular Question, t-test, Talkative, Promoter, Nice Answer, Benefactor, Nice Answer, Enlightened, Nice Answer, Nice Answer, Enlightened, Nice Answer, Enlightened, Steward, categorical-data, distributions, Informed, Nice Answer, Nice Answer, data-transformation, Nice Answer, Nice Answer, Steward, Revival, multicollinearity, model-selection, Nice Answer, Enlightened, Good Answer, Guru, probability, Nice Answer, mean, Nice Question, self-study, Nice Answer, modeling, Steward, Nice Answer, Nice Answer, Good Answer, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Enlightened, Curious, Nice Question, Nice Answer, Yearling, Nice Answer, Nice Answer, Enlightened, Inquisitive, Great Answer, Nice Answer, Good Answer, Refiner, Explainer, Sheriff, Nice Answer, Taxonomist, Nice Question, r, normal-distribution, Nice Answer	Teacher, Supporter, Organizer, Student, Scholar, Quorum, Editor, Commentator, Nice Answer, Autobiographer, Necromancer, Necromancer, Yearling, Nice Answer, Critic, Enthusiast, Nice Answer, Nice Answer, Necromancer, Good Answer, Nice Question, regression, Nice Answer, Popular Question, Civic Duty, Nice Answer, Nice Answer, Necromancer, Sportsmanship, Nice Answer, Nice Answer, Necromancer, Citizen Patrol, Necromancer, Nice Answer, Cleanup, Yearling, r, Nice Answer, Strunk & White, Custodian, logistic, Custodian, Revival, Custodian, Custodian, Analytical, Suffrage, Vox Populi, Tag Editor, Excavator, Custodian, Nice Question, correlation, Pundit, Nice Answer, Nice Answer, Custodian, Convention, Reviewer, Necromancer, Generalist, Fanatic, Good Question, Electorate, Proofreader, Custodian, Nice Answer, Reviewer, Reviewer, regression, Nice Answer, Mortarboard, Nice Answer, Copy Editor, multiple-regression, Good Answer, Nice Answer, Necromancer, statistical-significance, hypothesis-testing, Disciplined, Popular Question, Nice Question, Yearling, Nice Answer, Enlightened, Nice Answer, Popular Question, Nice Answer, Enlightened, Caucus, interaction, Constituent, anova, Popular Question, t-test, Talkative, Promoter, Nice Answer, Benefactor, Nice Answer, Enlightened, Nice Answer, Nice Answer, Enlightened, Nice Answer, Enlightened, Steward, categorical-data, distributions, Informed, Nice Answer, Nice Answer, data-transformation, Nice Answer, Nice Answer, Steward, Revival, multicollinearity, model-selection, Nice Answer, Enlightened, Good Answer, Guru, probability, Nice Answer, mean, Nice Question, self-study, Nice Answer, modeling, Steward, Nice Answer, Nice Answer, Good Answer, Nice Answer, Nice Answer, Nice Answer, Nice Answer, Enlightened, Curious, Nice Question, Nice Answer, Yearling, Nice Answer, Nice Answer, Enlightened, Inquisitive, Great Answer, Nice Answer, Good Answer, Refiner, Explainer, Sheriff, Nice Answer, Taxonomist, Nice Question, r, normal-distribution, Nice Answer	survival,simpsons-paradox,circular-statistics,time-series,panel-data,logistic,sas,margi distribution,multilevel analysis,unevenly-spaced-time-series,terminology,com power,ordered-logit,genetics,autocorr distribution,categorica data,mixed-model,interpolation,ge linear-model,multicollinearit analysis,missing-data,multiple-imputation,procrustes analysis,ranking,group differences,paradox,va to-the-mean,stratification,reg selection,linear-model,likelihood,mode binomial-distribution,nonparam test,pca,references,clu model,cross-validation,train,trend, for-trend,overfitting,repea measures,non-independent,simulatio normal-distribution,distributi statistics,kurtosis,teac moments,meta-analysis,splines,mars,r regression,intraclass-correlation,classificati testing,non-inferiority,matching,cl trials,random-allocation,fractional-polynomial,observatio study,number-needed-to-treat,quantiles,equival for-a-variable,censoring,dat visualization,proportio size,t-test,multiple-regression,quantile-regression,assumption

24 How many questions received no answers (accepted or unaccepted)?

In `Posts` table we count all rows of questions (`PostTypeId = 1`) which has no answers (`AnswerCount = 0`). 66970 are the number of unanswered questions.

```
SELECT COUNT(*) AS NoAnswers
FROM Posts
WHERE PostTypeId = 1 AND AnswerCount = 0;
```

Table 24: 1 records

NoAnswers
66970

24.1 How many questions had no accepted answer?

similarly to the query above, we count questions in `Posts`. However, this time we also filter on `AcceptedAnswerId = ""` to make sure there is none and also `AnswerCount > 0` to remove rows that has both no accepted or unaccepted answer. From this we can derive that there are about as many unanswered questions as there are questions with no accepted answer (given other answers)

```
SELECT COUNT(*) AS NumNoAcceptedAnswer
FROM Posts
WHERE PostTypeId = 1 AND AnswerCount > 0 AND AcceptedAnswerId = "";
```

Table 25: 1 records

NumNoAcceptedAnswer
69395

25 What is the distribution of answers per posted question?

Straightforward, we find the `AnswerCount` of questions (`PostTypeId = 1`) in the `Posts` table. We can see in the summary that most questions receive one answer.

```
SELECT AnswerCount
FROM Posts
WHERE PostTypeId = 1;
```

```
summary(df$AnswerCount)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##   0.0000   0.0000   1.0000   0.9685   1.0000 153.0000
```

26 What is the length of time for a question to receive an answer? to obtaining an accepted answer?

We start wit `AnswerDate` that maps each question to the date when its accepted answer was posted. Then, another CTE named `Dates` is created to compile the dates when each question was posted, its first answer was posted, and its accepted answer was posted. Finally, the main SELECT statement calculates the average time (in minutes), as suggested by LearnSQL (2023), between when each question was posted and when it received its first answer and between when each question was posted and when it received an accepted answer. We see that the accepted answer is often the first answer as well.

```
WITH AnswerDate AS (
  SELECT p.Id, pp.CreationDate AS AcceptedAnswerDate
  FROM Posts p
  INNER JOIN Posts pp ON p.AcceptedAnswerId = pp.Id
  WHERE p.PostTypeId = 1
  ),

Dates AS (
  SELECT
    pp.CreationDate AS QuestionPostedDate,
    MIN(p.CreationDate) AS FirstAnswer,
    ad.AcceptedAnswerDate
  FROM Posts p
  INNER JOIN Posts pp ON p.ParentId = pp.Id
  INNER JOIN AnswerDate ad ON pp.Id = ad.Id
  WHERE p.PostTypeId = 2
  GROUP BY pp.Id
  )

SELECT
  ROUND((julianday(FirstAnswer) - julianday(QuestionPostedDate)) * 24 *  60) as MinutesFirstAnswer,
  ROUND((julianday(AcceptedAnswerDate) - julianday(QuestionPostedDate)) * 24 * 60) as MinutesAcceptedAns
FROM Dates;
```

Table 26: Displaying records 1 - 10

MinutesFirstAnswer	MinutesAcceptedAnswer
8	8
12	30
1	1
138	143
3	8
202	44082
21752	21752
4	4
50	50
2	2

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

The query begins by creating `AnswerRank` where ChatGPT (2023) suggested to assign a rank to each answer based on the order of their posting for each question. Then, another CTE named `AcceptedAnswerRank` is created that maps each question with the rank of its accepted answer. Finally, we calculate the average rank of the accepted answers to get the number of answers before the accepted one.

As mentioned above, we confirm that most of the time, the first answer is the accepted answer.

```
WITH AnswerRank AS (
    SELECT
        ParentId as QuestionId,
        Id as AnswerId,
        CreationDate as AnswerCreationDate,
        ROW_NUMBER() OVER(PARTITION BY ParentId ORDER BY CreationDate) as AnswerRank
    FROM Posts
    WHERE PostTypeId = 2
),

AcceptedAnswerRank AS (
    SELECT
        a.QuestionId,
        a.AnswerRank,
        p.AcceptedAnswerId
    FROM AnswerRank a
    INNER JOIN Posts p ON a.QuestionId = p.Id
    WHERE p.AcceptedAnswerId = a.AnswerId
)

SELECT
    AVG(AnswerRank) as AverageAnswersBeforeAccepted
FROM AcceptedAnswerRank;
```

Table 27: 1 records

AverageAnswersBeforeAccepted
1.183607

28 Sources

- BMN. (2013). Answer to 'Using Sql Count in a Case Statement', Stack Overflow, Available Online: https://stackoverflow.com/a/17975288 [Accessed 18 May 2023]
- b8b8j. (2010). Answer to 'Query to Count Words SQLite 3', Stack Overflow, Available Online: https://stackoverflow.com/a/3293913 [Accessed 18 May 2023]
- ChatGPT. (2023). Available Online: https://chat.openai.com/ [Accessed 18 May 2023]
- Community Wiki. (2023). Answer to 'Database Schema Documentation for the Public Data Dump and SEDE', Meta Stack Exchange, Available Online: https://meta.stackexchange.com/a/2678 [Accessed 18 May 2023]

- Ian. (2021). How to Extract the Day, Month, and Year from a Date in SQLite, Available Online: https://database.guide/how-to-extract-the-day-month-and-year-from-a-date-in-sqlite/ [Accessed 21 May 2023]
- LearnSQL. (2023). How to Calculate the Difference Between Two Timestamps in SQLite, LearnSQL.Com, Available Online: https://learnsql.com/cookbook/how-to-calculate-the-difference-between-two-timestamps-in-sqlite/ [Accessed 21 May 2023]
- Lustig, L. (2010). Answer to 'Group Different Rows in One by Combining Strings', Stack Overflow, Available Online: https://stackoverflow.com/a/3926380 [Accessed 20 May 2023]