SQL in R

Code

Mark Gallo

- 1. Getting to know the data
- 2. Seeing the Devil
- 3. The Trial
- 4. Tortured Truth (Bonus)

Hide

```
library(RMySQL)
```

```
Loading required package: DBI
```

Hide

```
library(dbplyr)
```

```
Registered S3 method overwritten by 'dplyr':
  method           from
  print.rowwise_df
Registered S3 methods overwritten by 'dbplyr':
  method         from
  print.tbl_lazy
  print.tbl_sql
```

Code

```
<MySQLConnection:0,0>
```

Hide

```
DESCRIBE accused
```

Field <chr>	Type <chr>	Null <chr>	Key <chr>	Default <chr>	Extra <chr>
row_names	text	YES		*NA*	
accusedref	text	YES		*NA*	
accusedsystemid	text	YES		*NA*	
accusedid	double	YES		*NA*	
firstname	text	YES		*NA*	
lastname	text	YES		*NA*	
m_firstname	text	YES		*NA*	

Field <chr>	Type <chr>	Null <chr>	Key <chr>	Default <chr>	Extra <chr>
m_surname	text	YES		*NA*	
alias	text	YES		*NA*	
patronymic	text	YES		*NA*	

1-10 of 32 rows Previous **1** 2 3 4 Next

1. Getting to know the data

A. Show a list of the tables included in the database.

Hide

```
SHOW TABLES
```

Tables_in_witchcraft <chr>
accused
accused_family
appeal
calendarcustom
case_person
commission
complaint
confession
counterstrategy
demonicpact

1-10 of 37 rows Previous **1** 2 3 4 Next

B. Display the column names for the table accused.

Hide

```
select distinct COLUMN_NAME
from INFORMATION_SCHEMA.COLUMNS
where TABLE_NAME='accused'
```

COLUMN_NAME <chr>

COLUMN_NAME <chr>
row_names
accusedref
accusedsystemid
accusedid
firstname
lastname
m_firstname
m_surname
alias
patronymic

1-10 of 32 rows Previous **1** 2 3 4 Next

C.How many people are included in the accused table?

> 3217 People

Hide

```
SELECT DISTINCT Count(firstname), Count(lastname)
from accused
```

Count(firstname) <dbl>	Count(lastname) <dbl>
3217	3217

1 row

D. Display the columns firstname, sex, and age for 5 cases in the accused table.

Hide

```
SELECT DISTINCT firstname, sex, age
FROM accused
WHERE age > 70
```

firstname <chr>	sex <chr>	age <dbl>
Alexander	Male	75
Unknown	Female	100

firstname <chr>	sex <chr>	age <dbl>
Suna	Female	75
Marion	Female	80
Beatrix	Female	84

5 rows

E. Looks like the age is missing for some observations. Count the number of nonmissing values for age in the data.

> 166 individuals where age is provided

Hide

```
select count(age)
from accused
```

count(age) <dbl>
166

1 row

F.Show a list of unique occupations.

Hide

```
SELECT DISTINCT occupation
FROM accused
```

occupation <chr>
NA
Servant
Vagabond
Weaver
Midwife
Tailor
Messenger
Brewster
Smith

occupation <chr>
Minister

1-10 of 32 rows Previous **1** 2 3 4 Next

2. Seeing the Devil

A.List the unique devil_types in the data.

Hide

```
SELECT DISTINCT devil_type
FROM devilappearance
```

devil_type <chr>
Male
Female Fairy
Male Fairy
Animal Devil
Spirit
Ghost
Other Demon
Female
NA
Fairy

1-10 of 16 rows Previous **1** 2 Next

B. There is also a little description of the type of the devil sighting in the devil_text column. How many of the sightings mention the word “black” in the description?

121 Mentions

Hide

```
SELECT DISTINCT count(devil_text)
FROM devilappearance
WHERE devil_text LIKE "%black%"
```

count(devil_text) <dbl>
121

1 row

C. What proportion of the devils (in devil_type) are male?

63.13% Including NA, where NA is considered Not Male

Hide

```
SELECT (devil_type)
FROM devilappearance
```

devil_type <chr>
Male
Male
Male
Male
Male
Male
Male
Male
Female Fairy
Male

1-10 of 396 rows Previous **1** 2 3 4 5 6 … 40 Next

Hide

```
SELECT count(devil_type)
FROM devilappearance
WHERE devil_type LIKE "male"
```

count(devil_type) <dbl>
250

1 row

Hide

```
250/396
```

```
[1] 0.6313131
```

3. The Trial

A. What are the average and maximum numbers of male and female accusers?

Averages include when there were 0 accusers

Male

Max: 48

Avg: 0.418

Female

Max: 27

Avg: 0.256

Hide

```
SELECT male_accusers
from trial
ORDER BY male_accusers DESC
```

male_accusers
<dbl>

male_accusers <dbl>
48
43
25
25
25
25
24
23
22
22

1-10 of 1,000 rows Previous **1** 2 3 4 5 6 … 100 Next

Hide

```
SELECT avg(male_accusers)
FROM trial
```

avg(male_accusers) <dbl>
0.4181988

1 row

Hide

```
SELECT female_accusers
from trial
ORDER BY female_accusers DESC
```

female_accusers <dbl>
27
22
19
19
18

female_accusers <dbl>
16
16
16
16
16

1-10 of 1,000 rows Previous **1** 2 3 4 5 6 … 100 Next

Hide

```
SELECT AVG(female_accusers)
FROM trial
```

AVG(female_accusers) <dbl>
0.2559227

1 row

B. Count the number of sentences by sentence type. List them in a table (in descending order), excluding missing values. Rename the column headings to something sensible.

Hide

```
SELECT Sentence, Count(*) AS "Number of Sentences"
FROM trial
GROUP BY sentence
ORDER BY count(sentence) DESC
```

Sentence <chr>	Number of Sentences <dbl>
Execution	205
Released	52
Banishment	27
Declared Fugitive	11
Excommunicated	6
Put to the horn	2
Hang	1
Branded	1

Sentence <chr>	Number of Sentences <dbl>
Prison	1
Public Humiliation	1

1-10 of 11 rows Previous **1** 2 Next

C. Do the number of accusers matter for the verdict? Compare the average number of accusers by the type of verdict. Again make sure the table is sorted and the headings make sense.

Hide

```
SELECT Verdict, avg(male_accusers) AS "Average Number of Male Accursers", avg(female_accusers) A
S "Average Number of Female Accursers"
FROM trial
GROUP BY verdict
ORDER BY avg(male_accusers) DESC
```

Verdict <chr>	Average Number of Male Accursers <dbl>	Average Number of Female Accursers <dbl>
Guilty	3.4893617	2.31355932
Not Guilty	3.0681818	1.46511628
Not Proven	0.4545455	0.54545455
Half Guilty	0.1428571	0.14285714
NA	0.1308379	0.07042254

5 rows

4. Tortured Truth (Bonus)

A. Left join the trial and confession tables. For what share of trials does the database record confessions? Create a results table with the number of all trials, the number of confessions, and the share of trials with confessions recorded.

35.72% of trials have confessions on record

Hide

```
SELECT
  count(trial.trialid) AS "Trial ID",
  count(confession.confessionid) AS "Confession ID",
  ROUND((count(confession.confessionid) / count(trial.trialid) * 100),1) AS "Confession Share
 (%)"

FROM trial
LEFT JOIN confession
ON trial.trialid = confession.confessionid;
```

Trial ID <dbl>	Confession ID <dbl>	Confession Share (%) <dbl>
3522	1258	35.7

1 row

B. Only a small number of trials have records of torture. Is there a higher share of confessions among trials with records of torture than trials without such record? Hint: You will need to merge on the confession table.

Hide

```
SELECT
  count(trial.trialid) AS "Trial ID",
  count(confession.confessionid) AS "Confession ID",
  ROUND((count(confession.confessionid) / count(trial.trialid) * 100),1) AS "Confession Share
 (%)"

FROM trial
LEFT JOIN confession
ON trial.trialid = confession.confessionid;
```

Hide

```
describe confession
```

Field <chr>	Type <chr>	Null <chr>	Key <chr>	Default <chr>	Extra <chr>
row_names	text	YES		*NA*	
confessionref	text	YES		*NA*	
confessionsystemid	text	YES		*NA*	
confessionid	double	YES		*NA*	
trialref	text	YES		*NA*	
centraltrialconfession	tinyint(4)	YES		*NA*	
confessionrec	tinyint(4)	YES		*NA*	

Field <chr>	Type <chr>	Null <chr>	Key <chr>	Default <chr>	Extra <chr>
confessiondate	text	YES		*NA*	
confessiondate_as_date	text	YES		*NA*	
confessionplace	text	YES		*NA*	

1-10 of 19 rows Previous **1** 2 Next

Hide

```
select
  (trial.pretrialnotes),
  (confession.centraltrialconfession    )
from trial
Join confession
Where trial.pretrialnotes Like "%torture%_%was%_%used%"
```

pretrialnotes <chr>
Commission for trial issued but suspect had already been imprisoned. The commission specified that no torture v Text of the commission in JC10/4
Commission to try and judge the accused although she was already held in prison. PC stress that no torture to be to JC26/50 torture was used.
Commission to try the accused but PC stress no torture to be used. Suspect already in prison. Torture was used JC26/50.
Accused appears to have confessed to some degree of witchcraft but no details. PC stressed to the commissione confession had to contain detail about pact and renunciation of baptism etc, no torture was to be used and the ac mentally competent before the death sentence could be used.
Accused seems to have confessed to some aspect of witchcraft but no details. PC stressed to commissioners tha pact and renunciation etc could only be used if no torture was used and the accused was legally and mentally co penalty could only be used in these circumstances.
Commission for trial issued after some confession but council wanted more detail about her confession. No tortu and the commissioners needed to show that she was not legally or mentally incompetent.
Commission for trial issued but suspect had already been imprisoned. The commission specified that no torture v Text of the commission in JC10/4
Commission to try and judge the accused although she was already held in prison. PC stress that no torture to be to JC26/50 torture was used.
Commission to try the accused but PC stress no torture to be used. Suspect already in prison. Torture was used JC26/50.
Accused appears to have confessed to some degree of witchcraft but no details. PC stressed to the commissione confession had to contain detail about pact and renunciation of baptism etc, no torture was to be used and the ac mentally competent before the death sentence could be used.

1-10 of 1,000 rows | 1-1 of 2 columns Previous **1** 2 3 4 5 6 ... 100 Next