HW 10

Kendall Loh

12/3/2018

QMSS G5072 Homework 10

Setting up libraries

```
library(odbc)
library(DBI)
library(dbplyr)
library(magrittr)
library(dbplyr)
library(dplyr)
```

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

```
## The following objects are masked from 'package:dbplyr':
##
##     ident, sql
```

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

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

Setting up Connection

```
# Connection
witch_con <- dbConnect(RMySQL::MySQL(),
                  user='student',
                  password='mds-is-fun',
                  dbname='witchcraft',
                  host='tbrambor.cczjcoy6i2jn.us-east-1.rds.amazonaws.com',
                  port = 3306)
```

1. Getting to know the data

1a. Show a list of the tables included in the database.

```
SHOW TABLES
```

Table 1: Displaying records 1 - 10

Tables_in_witchcraft
accused
accused_family

Tables_in_witchcraft
appeal
calendarcustom
case_person
commission
complaint
confession
counterstrategy
demonicpact

1b. Display the column names for the table accused.

```
DESCRIBE accused
```

Table 2: Displaying records 1 - 10

Field	Type	Null	Key	Default	Extra
row_names	text	YES		NA	
accusedref	text	YES		NA	
accusedsystemid	text	YES		NA	
accusedid	bigint(20)	YES		NA	
firstname	text	YES		NA	
lastname	text	YES		NA	
m_firstname	text	YES		NA	
m_surname	text	YES		NA	
alias	text	YES		NA	
patronymic	text	YES		NA	

1c. How many people are included in the accused table?

```
SELECT COUNT(lastname)
FROM accused
```

Table 3: 1 records

COUNT(lastname)
3217

1d. Display the columns firstname, sex, and age for 5 cases in the accused table.

```
SELECT firstname, sex, age
FROM accused
LIMIT 5
```

Table 4: 5 records

firstname	sex	age
Mareon	Female	NA
Thom	Male	NA
Christian	Female	NA

firstname	sex	age
Janet	Female	NA
Agnes	Female	NA

1e. Looks like the age is missing for some observations. Count the number of nonmissing values for age in the data.

```
SELECT COUNT(age)
FROM accused
```

Table 5: 1 records

COUNT(age)
166

1f. how a list of unique occupations.

```
SELECT DISTINCT(occupation)
FROM accused
```

Table 6: Displaying records 1 - 10

occupation
NA
Servant
Vagabond
Weaver
Midwife
Tailor
Messenger
Brewster
Smith
Minister

2. Seeing the Devil

2a. Let's look at some appearances of the devil in the devilappearance table.

```
DESCRIBE devilappearance
```

Table 7: Displaying records 1 - 10

Field	Type	Null	Key	Default	Extra
row_names	text	YES		NA	
devilref	text	YES		NA	
devilsystemid	text	YES		NA	
devilid	bigint(20)	YES		NA	
caseref	text	YES		NA	
devil_type	text	YES		NA	

Field	Type	Null	Key	Default	Extra
devil_text	text	YES		NA	
createdby	text	YES		NA	
createdate	text	YES		NA	
lastupdatedby	text	YES		NA	

2b. List the unique devil_types in the data.

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

Table 8: Displaying records 1 - 10

devil_type
Male
Female Fairy
Male Fairy
Animal Devil
Spirit
Ghost
Other Demon
Female
NA
Fairy

2c. 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?

```
SELECT COUNT(*)
FROM devilappearance
WHERE devil_text LIKE '%black%'
```

Table 9: 1 records

COUNT(*)
121

2d. What proportion of the devils (in devil_type) are male?

```
SELECT
 (SUM(devil_type = 'MALE')/COUNT(*)) * 100 as Proportion
FROM devilappearance
```

Table 10: 1 records

Proportion
63.1313

3. The trial

Let's take a look at the information on the trial.

```
DESCRIBE trial
```

Table 11: Displaying records 1 - 10

Field	Type	Null	Key	Default	Extra
row_names	text	YES		NA	
trialref	text	YES		NA	
trialid	bigint(20)	YES		NA	
trialsystemid	text	YES		NA	
caseref	text	YES		NA	
trialtype	bigint(20)	YES		NA	
trial_settlement	text	YES		NA	
trial_parish	text	YES		NA	
trial_presbytery	text	YES		NA	
trial_county	text	YES		NA	

3a. What are the average and maximum numbers of male and female accusers?

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

Table 12: 1 records

AVG (female_accusers)
0.2559

```
SELECT MAX(female_accusers)
FROM trial
```

Table 13: 1 records

MAX(female_accusers)
27

```
SELECT AVG (male_accusers)
FROM trial
```

Table 14: 1 records

AVG (male_accusers)
0.4182

```
SELECT MAX(male_accusers)
```

```
FROM trial
```

Table 15: 1 records

MAX(male_accusers)
48

3b. 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.

```
SELECT sentence AS "Sentence Type", COUNT(*) AS "Sentence Volume"
FROM trial
GROUP BY sentence
ORDER BY COUNT(*) DESC
```

Table 16: Displaying records 1 - 10

Sentence Type	Sentence Volume
NA	2904
Execution	205
Released	52
Banishment	27
Declared Fugitive	11
Excommunicated	6
Put to the horn	2
Hang	1
Branded	1
Prison	1

3c. 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.

```
SELECT verdict AS "Verdict", AVG(female_accusers) AS "Mean Number Female Accusers", AVG(male_accusers) A
FROM trial
GROUP BY verdict
ORDER BY AVG(female_accusers) DESC
```

Table 17: 5 records

Verdict	Mean Number Female Accusers	Mean Number of Male Accusers
Guilty	2.3136	3.4894
Not Guilty	1.4651	3.0682
Not Proven	0.5455	0.4545
Half Guilty	0.1429	0.1429
NA	0.0704	0.1308

4. Tortured Truth (Bonus)

Note: This part is optional. We spent little time on joins in lecture, so I encourage you to try it but feel free to skip.

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.

```
SHOW TABLES
```

Table 18: Displaying records 1 - 10

Tables_in_witchcraft
accused
accused_family
appeal
calendarcustom
case_person
commission
complaint
confession
counterstrategy
demonicpact

```
DESCRIBE trial
```

Table 19: Displaying records 1 - 10

Field	Type	Null	Key	Default	Extra
row_names	text	YES		NA	
trialref	text	YES		NA	
trialid	bigint(20)	YES		NA	
trialsystemid	text	YES		NA	
caseref	text	YES		NA	
trialtype	bigint(20)	YES		NA	
trial_settlement	text	YES		NA	
trial_parish	text	YES		NA	
trial_presbytery	text	YES		NA	
trial_county	text	YES		NA	

```
SELECT
  (COUNT(trialid)) AS "Trial Count",
  (COUNT(confessionid)) AS "Confession Count",
  (COUNT(confessionid)/COUNT(trialid)) * 100 as Proportion
FROM trial
LEFT JOIN confession
ON trial.trialref=confession.trialref
```

Table 20: 1 records

Trial Count	Confession Count	Proportion
3396	941	27.7091

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.

110 tortured 3211 trials 941 confessions