

What are the Cowboyest Names in Switzerland?

Data Management and Integration Assignment

Marius Furter and Morley J Weston

2022-11-23

```
library(tidyverse)
library(DBI)
library(knitr)
```

Create the database

```
conn <- dbConnect(RSQLite::SQLite(), "cowboys.db")
```

Population by Commune

The data was gathered from the Swiss federal geodata portal at data.geo.admin.ch, and converted from an ESRI shapefile into a CSV in QGIS.

```
communes <- read_csv("input_data/gemeinde.csv")
communes <- communes %>%
  filter(!is.na(EINWOHNERZ)) %>%
  select(BFS_NUMMER, EINWOHNERZ, NAME)

communes %>% sample_n(10) %>% kable()
```

BFS_NUMMER	EINWOHNERZ	NAME
7002	5330	Triesen
3713	79	Ferrera
743	6943	Nidau
500	429	Treiten
700	7348	Moutier
6806	560	Vendlincourt
5742	376	Agiez
70	568	Wasterkingen
5186	477	Grancia
2962	471	Hemishofen

We created a table for the communes with the unique ID as a primary key ...

```
dbExecute(conn, "
  create table if not exists communes (
    commune_id int primary key,
    population int,
```

```

    name text
  )
")

## [1] 0

for (i in 1:nrow(communes)) {
  row = communes[i,]
  dbExecute(conn, "
    insert into communes (commune_id, population, name)
    values (?, ?, ?)",
    params = c(row$BFS_NUMMER, row$EINWOHNERZ, row$NAME))
}

dbGetQuery(conn, "select * from communes limit 10") %>% kable()

```

commune_id	population	name
3762	4624	Scuol
1631	9480	Glarus Süd
3746	1506	Zernez
3543	2377	Surses
6037	10329	Val de Bagnes
3851	10832	Davos
3792	1556	Bregaglia
6252	2742	Anniviers
6300	5820	Zermatt
784	1072	Innertkirchen

Cattle population per commune

```

cattle_population <- read_delim("input_data/cattle-map-commune.csv",
  delim = ";", escape_double = FALSE, trim_ws = TRUE)
cattle_population <- cattle_population %>%
  select(MunicipalityNumber, count)

cattle_population %>% sample_n(10) %>% kable()

```

MunicipalityNumber	count
5637	47
305	661
6134	182
5566	1572
5146	0
4	1285
5403	191
2455	113
3722	244
738	165

TODO: 1. alter table communes to add a cattle_population column, 2. update communes cattle_population where commune_id is equal

www.sqlitetutorial.net is a good resource

Most common family names by commune

```
family_names <-  
  read_delim(  
    "input_data/family-names-commune.csv",  
    delim = ";",  
    escape_double = FALSE,  
    trim_ws = TRUE  
  )  
family_names <- family_names %>%  
  filter(TIME_PERIOD == 2021) %>%  
  filter(RANG_GDE <= 3) %>%  
  select(LASTNAME, GDENR, RANG_GDE, VALUE)  
  
family_names %>% sample_n(10) %>% kable()
```

LASTNAME	GDENR	RANG_GDE	VALUE
Ferreira	2228	2	56
Christen	4233	3	13
Moser	607	1	17
Schmid	4561	1	29
Bachmann	293	3	119
Gasser	1211	3	18
Charbon	5828	2	7
Fornera	5115	1	76
Popp	4421	2	23
Mulaj	4139	1	64

```
dbExecute(conn, "  
  create table if not exists family_names (  
    id integer primary key,  
    name text,  
    commune_id int,  
    name_count int,  
    rank int,  
    foreign key(commune_id) references communes(commune_id)  
  )  
")  
  
## [1] 0  
  
pb = txtProgressBar(min = 1, max = nrow(family_names), initial = 1)  
  
for (i in 1:nrow(family_names)) {  
  row = family_names[i,]  
  dbExecute(conn, "  
    insert into family_names (name, commune_id, name_count, rank)  
    values (?, ?, ?, ?)",  
    params = c(row$LASTNAME, row$GDENR, row$VALUE, row$RANG_GDE))  
}
```

```

if (i %% 100 == 0) {
  setTxtProgressBar(pb,i)}
}

```

```
## =====
```

```
close(pb)
```

```
dbGetQuery(conn, "select * from family_names limit 10") %>% kable()
```

	id	name	commune_id	name_count	rank
	1	Frei	1	14	1
	2	Meier	1	14	1
	3	Weiss	1	14	1
	4	Schneebeli	2	77	1
	5	Müller	2	68	2
	6	Krasniqi	2	66	3
	7	Müller	3	54	1
	8	Meier	3	50	2
	9	Huber	3	32	3
	10	Huber	4	40	1

Most common cattle names by commune

```

cattle_NamesFemaleCalves <- read_delim("input_data/cattle-NamesFemaleCalves.csv",
  delim = ";", escape_double = FALSE, trim_ws = TRUE) %>%
  mutate(sex="F")

```

```

cattle_NamesMaleCalves <- read_delim("input_data/cattle-NamesMaleCalves.csv",
  delim = ";", escape_double = FALSE, trim_ws = TRUE) %>%
  mutate(sex="M")

```

```

cattle_names <- bind_rows(cattle_NamesFemaleCalves, cattle_NamesMaleCalves)
rm(cattle_NamesFemaleCalves, cattle_NamesMaleCalves)

```

```

cattle_names <- cattle_names %>%
  filter(year==2022) %>%
  filter(OwnerLanguage != "__all__") %>%
  select(Name, count, Rank, sex, OwnerLanguage)

```

```
cattle_names %>% arrange(Rank, sex, OwnerLanguage) %>% head(10) %>% kable()
```

Name	count	Rank	sex	OwnerLanguage
Bella	674	1	F	de
Tulipe	88	1	F	fr
Luna	12	1	F	it
Max	633	1	M	de
Pvv	133	1	M	fr
Rambo	7	1	M	it
Sina	518	2	F	de
Bella	77	2	F	fr

Name	count	Rank	sex	OwnerLanguage
Gina	10	2	F	it
Leo	545	2	M	de

TODO make table for this