

Assignment 1 - The tidyverse

Adrian Werner

September 24, 2022 | 16:18:06 | CEST

```
library("tidyverse")
library(legislatoR)
library(lubridate)
```

Getting started with the Comparative Legislators Database

The Comparative Legislators Database (CLD) includes political, sociodemographic, career, online presence, public attention, and visual information for over 45,000 contemporary and historical politicians from ten countries (see the corresponding article for more information). It can be accessed via `legislatoR` - an R package that is available on both CRAN and GitHub.

Before you start with the following tasks, skim the tutorial to make yourself familiar with the database. You find it here.

For the following tasks, you will work with ONE of the legislatures. The legislature you'll work with depends on your first name:

Your first name starts with...	Legislature	Code
A-C	Austrian Nationalrat	<code>aut</code>
D-F	Canadian House of Commons	<code>can</code>
G-I	Czech Poslanecka Snemovna	<code>cze</code>
J-L	Spanish Congreso de los Diputados	<code>esp</code>
M-O	French Assemblée	<code>fra</code>
P-R	German Bundestag	<code>deu</code>
S-U	Irish Dail	<code>irl</code>
V-X	UK House of Commons	<code>gbr</code>
Y-Z	US House of Representatives	<code>usa_house</code>

The following tasks will be based on data from the `core` and `political` tables of the database.

Task 1 - Descriptives [8 points in total]

- a) What's the overall share of male legislators in the entire dataset? [1 point]

```
aut_house_core <- get_core(legislature = "aut")
glimpse(aut_house_core)
```

Rows: 1,923

Columns: 12

```
$ country    <chr> "AUT", "AUT", "AUT", "AUT", "AUT", "AUT", "AUT", "AUT", "AU~
$ pageid     <chr> "3041286", "377116", "3521961", "4945354", "880210", "39624~
$ wikidataid <chr> "Q2287337", "Q93503", "Q1908424", "Q1609919", "Q1596544", "~
$ wikititle  <chr> "Simon_Abram", "Friedrich_Adler_(Politiker)", "Mathias_Adle~
$ name       <chr> "Abram Simon", "Adler Friedrich", "Adlersflügel Mathias", "~
$ sex        <chr> "male", "male", "male", "male", "male", "male", "male", "ma~
$ ethnicity  <chr> "white", "white", NA, NA, NA, NA, NA, NA, NA, "white", NA, ~
$ religion    <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "catholicis~
$ birth      <dtm> 1871-04-03, 1879-07-09, 1868-01-17, 1884-03-19, 1878-11-24~
$ death      <dtm> 1940-02-29, 1960-01-02, 1933-09-01, 1947-08-05, 1953-12-10~
$ birthplace <chr> "47.2569,11.3861", "48.20833,16.37306", "48.20833,16.37306"~
$ deathplace <chr> "47.8,13.03333", "47.37861,8.54", "48.20833,16.37306", "48.~
```

```
table(aut_house_core$sex)/length(aut_house_core$sex)
```

```
female    male
0.1861674 0.8138326
```

- b) How many (both in absolute and relative terms) legislators died in the same place they were born in? [1 point]

```
aut_house_core %>%
  filter(aut_house_core$birthplace == aut_house_core$deathplace) %>%
  nrow()
```

```
[1] 253
```

```
aut_house_core %>%
  filter(aut_house_core$birthplace == aut_house_core$deathplace) %>%
  nrow()/nrow(aut_house_core)
```

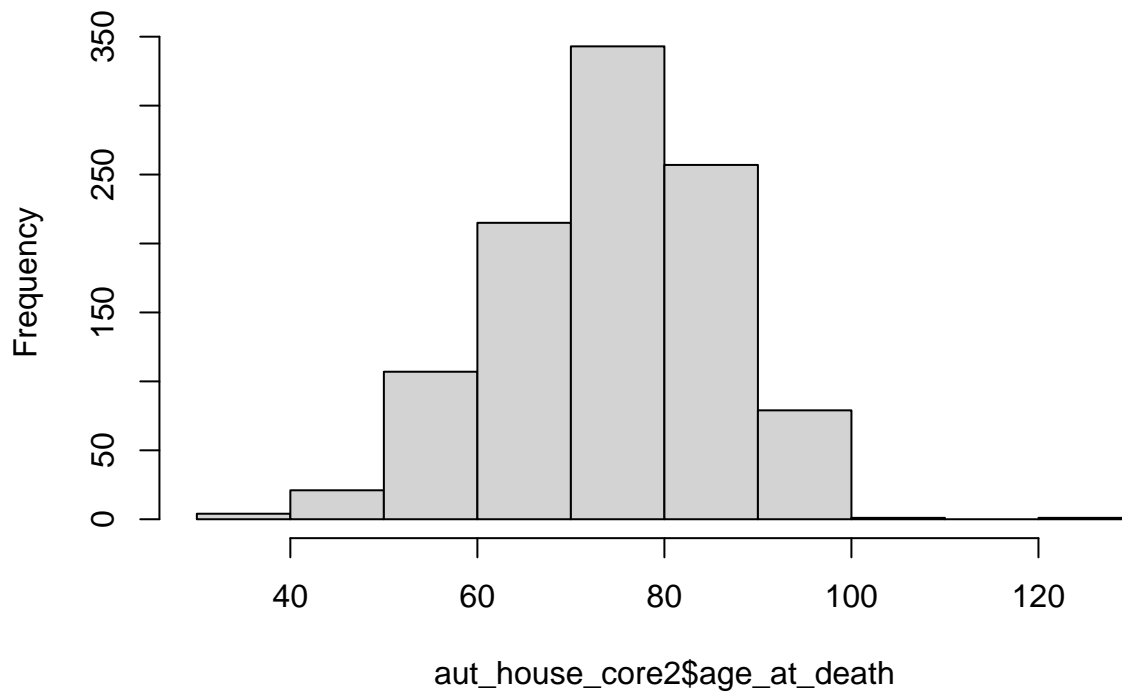
```
[1] 0.1315653
```

- c) Create a new variable `age_at_death` that reports the age at which legislators died. Then, plot the distribution of that variable for the entire dataset. [2 points]

```
aut_house_core2 <- aut_house_core %>%
  select(name, death, birth) %>%
  mutate(
    age_at_death = time_length(difftime(aut_house_core$death, aut_house_core$birth), "years")
  )

hist(aut_house_core2$age_at_death)
```

Histogram of aut_house_core2\$age_at_death



d) What is the most frequent birthday in your sample (e.g., “January 15”)? [2 points]

```
birth_fac <- as.factor(aut_house_core$birth)
names(which.max(table(birth_fac)))
```

```
[1] "1975-05-07"
```

e) Generate a table that provides the 5 legislators with the longest names (in terms of number of characters, ignoring whitespace). [2 points]

```
aut_house_core3 <- aut_house_core %>%
  select(name) %>%
  mutate(
    character_length = nchar(name)) %>%
  arrange(desc(character_length))
)

aut_house_core4 <- head(aut_house_core3, 5)

aut_house_core4
```

	name	character_length
698	Schauer-Schoberlechner Johannes	31
4668	Kaufmann-Bruckberger Elisabeth	30

4594	Cortolezis-Schlager Katharina	29
4866	Holzinger-Vogtenhuber Daniela	29
5074	Hoyos-Trauttmansdorff Douglas	29

Task 2 - Exploring trends in the data [8 points in total]

- a) Using data from all sessions, provide a plot that shows the share of male legislators by session! [3 points]

```
aut_house_pol <- get_political(legislature = "aut")

merged_df <- merge.data.frame(aut_house_core, aut_house_pol)

merged_df %>%
  mutate(gender = ifelse(merged_df$sex == "male", 1, 0)) %>%
  group_by(session) %>%
  summarise(perc_male = ((sum(gender)/length(gender))))
```

```
# A tibble: 27 x 2
  session perc_male
  <int>     <dbl>
1     1     0.941
2     2     0.951
3     3     0.966
4     4     0.945
5     5     0.944
6     6     0.942
7     7     0.938
8     8     0.944
9     9     0.939
10    10     0.940
# ... with 17 more rows
```

- b) Explore another relation of variables in the dataset. Your analysis should (i) use data from both the `core` and the `political` table and (ii) feature data from several sessions. Also, offer a brief description and interpretation of your findings! [5 points]

```
merged_df %>%
  mutate(death_no = ifelse(is.na(merged_df$death), 0, 1)) %>%
  group_by(party) %>%
  summarise(death_no = sum(death_no)/length(death_no)) %>%
  arrange(desc(death_no))
```

```
# A tibble: 19 x 2
  party    death_no
  <chr>      <dbl>
1 CsP         1
2 GdP         1
3 HB          1
```

4	KPÖ	1
5	LBd	1
6	NWB	1
7	SdP	1
8	VO	1
9	WdU	1
10	none	0.818
11	SPÖ	0.559
12	ÖVP	0.537
13	FPÖ	0.177
14	GRÜNE	0.0110
15	BZÖ	0
16	LIF	0
17	NEOS	0
18	PILZ	0
19	STRONACH	0

The table shows, in absolute terms, the death rate among party members in a descending order. Interestingly, there are a number of parties with a death rate of 100%. This is because they are members of parties that do not exist any longer. These parties are therefore not only formally but also in terms of their representatives vanished from the political landscape. Interestingly, even the biggest parties SPÖ (founded 1889) and ÖVP (founded 1945) are >50% “dead”, indicating that Austria is already a old democracy.
