btmembers: An R package to import data on all members of the Bundestag since 1949

Philippe Joly
Otto Suhr Institute of Political Science, Free University of Berlin

April 26, 2021

Abstract

The Bundestag distributes biographical and election data on all its members since 1949. This data, however, is only available in XML, a format social scientists might find difficult to work with. This paper introduces a tool to make the Bundestag open data more accessible: the btmembers R package. btmembers downloads the XML file on members of the Bundestag, converts it to a data frame, and recodes some of the variables. The generated dataset contains more than 11,000 observations for more than 4,000 members of the Bundestag. With its tabular format, the dataset allows users to easily examine the evolution of the composition of the parliament with regards to gender, age, occupation, and other characteristics. The package is a useful resource for researchers, journalists, teachers, and the broader public.

Keywords: Bundestag, parliament, open data, political representation, R

The Bundestag website provides useful resources such as minutes of the plenary sessions, roll call votes, and data on elected members. Information on members of the Bundestag, however, can be difficult to extract since it is stored in an XML file. Unlike usual rectangular datasets, XML files have a tree-like structure. They can organize data in an arbitrary number of dimensions. This format is not well-suited for comparative analyses. In this short paper, I would like to introduce a tool to make the Bundestag open data accessible to a broader public: the btmembers R package. btmembers downloads the XML file on members of the Bundestag, converts it to a data frame, and recodes some of the variables. The generated dataset contains more than 11,000 observations for more than 4,000 members of the Bundestag.

1 A look at the original data

To illustrate the challenges of working with the original XML file provided by the Bundestag, let us have a look at the data for Elisabeth Schwarzhaupt (CDU). Schwarzhaupt was Federal Minister of Health from 1961 to 1966 and the first woman to hold a ministerial position in Germany.

```xml
<MDB>
    <ID>11002129</ID>
    <NAMEN>
      <NAME>
        <NACHNAME>Schwarzhaupt</NACHNAME>
        <VORNAME>Elisabeth</VORNAME>
        <ORTSZUSATZ/>
        <ADEL/>
        <PRAEFIX/>
        <ANREDE_TITEL>Dr.</ANREDE_TITEL>
        <AKAD_TITEL>Dr.</AKAD_TITEL>
        <HISTORIE_VON>06.10.1953</HISTORIE_VON>
        <HISTORIE_BIS/>
      </NAME>
    </NAMEN>
    <BIOGRAFISCHE_ANGABEN>
      <GEBURTSDATUM>07.01.1901</GEBURTSDATUM>
      <GEBURTSORT>Frankfurt/Main</GEBURTSORT>
      <GEBURTSLAND/>
      <STERBEDATUM>29.10.1986</STERBEDATUM>
      <GESCHLECHT>weiblich</GESCHLECHT>
      <FAMILIENSTAND>keine Angaben</FAMILIENSTAND>
      <RELIGION>evangelisch</RELIGION>
      <BERUF>Bundesminister für Gesundheitswesen, Oberkirchenrätin i. R.</BERUF>
      <PARTEI_KURZ>CDU</PARTEI_KURZ>
      <VITA_KURZ/>
      <VEROEFFENTLICHUNGSPFLICHTIGES/>
    </BIOGRAFISCHE_ANGABEN>
    <WAHLPERIODEN>
      <WAHLPERIODE>
        <WP>2</WP>
        <MDBWP_VON>06.10.1953</MDBWP_VON>
        <MDBWP_BIS>06.10.1957</MDBWP_BIS>
        <WKR_NUMMER/>
        <WKR_NAME/>
        <WKR_LAND/>
        <LISTE>HES</LISTE>
        <MANDATSART>Landesliste</MANDATSART>
        <INSTITUTIONEN>
          <INSTITUTION>
            <INSART_LANG>Fraktion/Gruppe</INSART_LANG>
            <INS_LANG>Fraktion der Christlich Demokratischen Union/Christlich - Sozialen Union</INS_LANG>
            <MDBINS_VON/>
            <MDBINS_BIS/>
            <FKT_LANG/>
            <FKTINS_VON/>
            <FKTINS_BIS/>
          </INSTITUTION>
        </INSTITUTIONEN>
      </WAHLPERIODE>
      <WAHLPERIODE>
        <WP>3</WP>
        <MDBWP_VON>15.10.1957</MDBWP_VON>
        <MDBWP_BIS>15.10.1961</MDBWP_BIS>
        <WKR_NUMMER>138</WKR_NUMMER>
        <WKR_NAME/>
        <WKR_LAND>HES</WKR_LAND>
        <LISTE/>
        <MANDATSART>Direktwahl</MANDATSART>
        <INSTITUTIONEN>
          <INSTITUTION>
            <INSART_LANG>Fraktion/Gruppe</INSART_LANG>
            <INS_LANG>Fraktion der Christlich Demokratischen Union/Christlich - Sozialen Union</INS_LANG>
            <MDBINS_VON/>
            <MDBINS_BIS/>
            <FKT_LANG/>
          <FKTINS_VON/>
          <FKTINS_BIS/>
        </INSTITUTION>
      </INSTITUTIONEN>
    </WAHLPERIODE>
    <WAHLPERIODE>
      <WP>4</WP>
      <MDBWP_VON>17.10.1961</MDBWP_VON>
      <MDBWP_BIS>17.10.1965</MDBWP_BIS>
      <WKR_NUMMER/>
      <WKR_NAME/>
      <WKR_LAND/>
      <LISTE>HES</LISTE>
      <MANDATSART>Landesliste</MANDATSART>
      <INSTITUTIONEN>
        <INSTITUTION>
          <INSART_LANG>Fraktion/Gruppe</INSART_LANG>
          <INS_LANG>Fraktion der Christlich Demokratischen Union/Christlich - Sozialen Union</INS_LANG>
          <MDBINS_VON/>
          <MDBINS_BIS/>
          <FKT_LANG/>
          <FKTINS_VON/>
          <FKTINS_BIS/>
        </INSTITUTION>
      </INSTITUTIONEN>
    </WAHLPERIODE>
    <WAHLPERIODE>
      <WP>5</WP>
      <MDBWP_VON>19.10.1965</MDBWP_VON>
      <MDBWP_BIS>19.10.1969</MDBWP_BIS>
      <WKR_NUMMER/>
      <WKR_NAME/>
      <WKR_LAND/>
      <LISTE>HES</LISTE>
      <MANDATSART>Landesliste</MANDATSART>
      <INSTITUTIONEN>
        <INSTITUTION>
          <INSART_LANG>Fraktion/Gruppe</INSART_LANG>
          <INS_LANG>Fraktion der Christlich Demokratischen Union/Christlich - Sozialen Union</INS_LANG>
          <MDBINS_VON>19.10.1965</MDBINS_VON>
          <MDBINS_BIS>19.10.1969</MDBINS_BIS>
          <FKT_LANG/>
          <FKTINS_VON/>
          <FKTINS_BIS/>
        </INSTITUTION>
      </INSTITUTIONEN>
    </WAHLPERIODE>
  </WAHLPERIODEN>
</MDB>
```

The file distributed by the Bundestag contains an XML-node named MDB (Mitglied des Deutschen Bundestages) for each member of the Bundestag. This node has four children: ID (id), NAMEN (names), BIOGRAFISCHE_ANGABEN (biographical information), and WAHLPERIODEN (parliamentary terms). All of these four children, except ID, have descendants, that is, data nested within different dimensions. In the example above, we see that Elisabeth Schwarzhaupt was born on January 7, 1901 in Frankfurt and died on October 20, 1986. She served four terms from 1953 to 1969.

Now, how do we combine information on multiple members? If we are not only interested in a specific member of the Bundestag but want to compare groups of members, we need a different data structure. I was faced with this problem while working on a chapter in an edited volume on the AfD[1] and started preparing R scripts to reshape the data. A few months later, I turned these scripts into a proper package to hopefully make the Bundestag open data a little more accessible. In the rest of this paper, I will introduce the btmembers R package, its applications, its remaining problems, and some potential improvements in the future.

2 Getting started with btmembers

The objective of btmembers is to import the file "Stammdaten aller Abgeordneten seit 1949 im XML-Format" and turn it into a tidy rectangular dataset. This involves both the mechanical operation of moving values into tabular cells and some choices about what information to keep or drop. The unit of analysis in the generated dataset is a member-term. In other words, each member can have multiple observations (multiple rows). Some variables are constant for each member (e.g., date of birth) while others vary over time (e.g., start date of a parliamentary term). The dataset preserves almost all of the original

[1] Schroeder, W., Weßels, B., & and Joly, P. (2019). Die AfD als Provokateur: Metamorphosen einer Partei zwischen Parlament und Bewegung. In W. Schroeder & B. Weßels (Eds.), *Smarte Spalter: Die AfD zwischen Bewegung und Parlament* (pp. 221-256). Bonn: Dietz.

data, except two elements:

- Only the most recent names of the member are preserved. Members can have multiple names, for example if they got married (or divorced).
- The “functions” of the member during a parliamentary term are not reported. Term variables are already nested in member variables. Incorporating the functions into the dataset would add yet another level of analysis since members can carry out multiple functions during the same term. This would make the structure of the dataset too complex. Furthermore, a look at the original data revealed that this variable was not coded systematically. It often omits functions carried out by the members. Information on functions is much more detailed for current than for previous members.

I will address some of the concerns related to the omission of these pieces of information at the end of this paper. The next sections describe how to import the data processed by btmembers.

2.1 Installation

btmembers is currently hosted on GitHub. I am considering submitting to CRAN at a later stage. For now, you can install the package using `devtools`.

```r
# install.packages("devtools")
devtools::install_github("jolyphil/btmembers")
```

2.2 Preloaded data

btmembers comes preloaded with the processed tabular dataset. The data is stored in an object called `members` and can be retrieved as follows:[2]

```r
library(btmembers)
members
```

[2] A CSV version of the dataset is also available on GitHub .

```
## # A tibble: 11,627 x 26
##    id       nachname  vorname adel  praefix anrede_titel akad_titel geburtsdatum
##    <chr>    <chr>     <chr>   <chr> <chr>   <chr>        <chr>      <date>
##  1 11000001 Abelein   Manfred <NA>  <NA>    Dr.          Prof. Dr.  1930-10-20
##  2 11000001 Abelein   Manfred <NA>  <NA>    Dr.          Prof. Dr.  1930-10-20
##  3 11000001 Abelein   Manfred <NA>  <NA>    Dr.          Prof. Dr.  1930-10-20
##  4 11000001 Abelein   Manfred <NA>  <NA>    Dr.          Prof. Dr.  1930-10-20
##  5 11000001 Abelein   Manfred <NA>  <NA>    Dr.          Prof. Dr.  1930-10-20
##  6 11000001 Abelein   Manfred <NA>  <NA>    Dr.          Prof. Dr.  1930-10-20
##  7 11000001 Abelein   Manfred <NA>  <NA>    Dr.          Prof. Dr.  1930-10-20
##  8 11000002 Achenbach Ernst   <NA>  <NA>    Dr.          Dr.        1909-04-09
##  9 11000002 Achenbach Ernst   <NA>  <NA>    Dr.          Dr.        1909-04-09
## 10 11000002 Achenbach Ernst   <NA>  <NA>    Dr.          Dr.        1909-04-09
## # ... with 11,617 more rows, and 18 more variables: geburtsort <chr>,
## #   geburtsland <chr>, sterbedatum <date>, geschlecht <chr>,
## #   familienstand <chr>, religion <chr>, beruf <chr>, partei_kurz <chr>,
## #   vita_kurz <chr>, veroeffentlichungspflichtiges <chr>, wp <int>,
## #   mdbwp_von <date>, mdbwp_bis <date>, wkr_nummer <int>, wkr_name <chr>,
## #   wkr_land <chr>, liste <chr>, mandatsart <chr>
```

2.3 Variables

The current version of the dataset (2021-03-12) contains 11627 observations from 4089 members. The data includes 26 variables (in German):[3]

[3] A codebook is available in the associated help page (`?members`) and on GitHub .

Variable	Type	Label
id	character	Identifikationsnummer
nachname	character	Nachname
vorname	character	Vorname
adel	character	Adelsprädikat
praefix	character	Namenspräfix
anrede_titel	character	Anrede-Titel
akad_titel	character	Akademischer Titel
geburtsdatum	Date	Geburtsdatum
geburtsort	character	Geburtsort
geburtsland	character	Geburtsland
sterbedatum	Date	Sterbedatum
geschlecht	character	Geschlecht
familienstand	character	Familienstand
religion	character	Religion
beruf	character	Beruf
partei_kurz	character	Parteizugehörigkeit, kurzform
vita_kurz	character	Kurzbiografie des Abgeordneten (nur aktuelle Wahlperiode)
veroeffentlichungspflichtiges	character	Veröffentlichungspflichtige Angaben (nur aktuelle Wahlperiode)
wp	integer	Nummer der Wahlperiode
mdbwp_von	Date	Beginn der Wahlperiodenzugehörigkeit
mdbwp_bis	Date	Ende der Wahlperiodenzugehörigkeit
wkr_nummer	integer	Nummer des Wahlkreises
wkr_name	character	Wahlkreisname
wkr_land	character	Bundesland des Wahlkreises
liste	character	Liste
mandatsart	character	Art des Mandates

2.4 Updating the data

One of the main advantage of btmembers over other available datasets is the easiness with which users can update the data. The Bundestag updates the XML file a few times every year.

The version of the dataset preloaded with your installation of btmembers is stored as an attribute of the `members` object. You can find the version loaded on your machine by proceeding as follows:

```
attr(members, "version")
```

```
## [1] "2021-03-12"
```

To check if a more recent version of the data is available on the bundestag website, simply call the following function:

```r
update_available()
```

```
## You currently have the most recent version of the data.
```

```
## [1] FALSE
```

If `update_available()` returns `TRUE`, you can import a more recent version of the dataset using `import_members()`. This function will follow the same procedure that generated the preloaded dataset. Importing and converting a new version of the dataset might take 2 to 3 minutes.

```r
members_new <- import_members()
```

Note that `members_new` is only stored in your global environment. You may save it on your disk using `saveRDS()`.

```r
saveRDS(members_new, "members_new.rds")
```

3 Applications

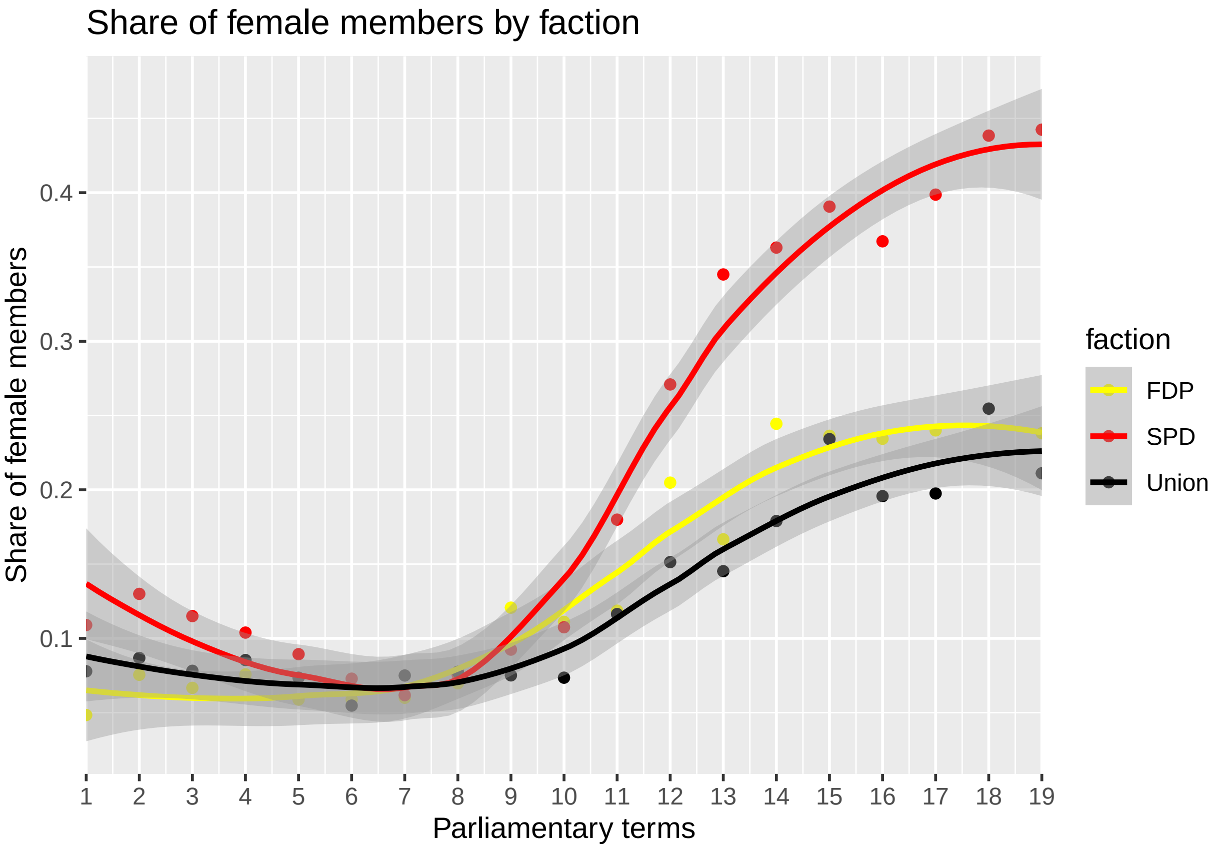
btmembers is particularly useful for research on political representation. It is also a great tool for teaching.

3.1 Research on political representation

Descriptive representation is the idea that the composition of parliaments (for example, in terms of gender, age, ethnicity, and social structure) should mirror the constituencies represented by elected members. With btmembers, it is easy to track the representation of certain groups in the Bundestag, for example, women. Below is an example showing how to examine the representation of women in different factions over time.

```r
library(dplyr)
share_women <- members %>%
  mutate(faction = case_when(
    partei_kurz == "CDU" | partei_kurz == "CSU" ~ "Union",
    partei_kurz == "SPD" ~ "SPD",
    partei_kurz == "FDP" ~ "FDP")) %>%
  filter(!is.na(faction)) %>%
  group_by(faction, wp) %>%
  summarize(women = mean(geschlecht == "weiblich"))
```

```r
library(ggplot2)
ggplot(data = share_women, aes(x = wp, y = women, color = faction)) +
  geom_point() +
  geom_smooth() +
  scale_x_continuous(expand = c(0,0), breaks = 1:19) +
  scale_color_manual(values=c("yellow", "red", "black")) +
  ggtitle("Share of female members by faction") +
  xlab("Parliamentary terms") +
  ylab("Share of female members")
```



The figure above shows that the share of female members in the FDP, SPD, and CDU/CSU factions remained very low—usually below 10 percent—until the end of the 1970s (8th parliamentary term). It then grew to reach about 20% in the FDP and CDU/CSU factions and over 40% in the SPD faction.

We can also have a look at changes in the occupational structure of members over time. The code below returns the ten most common occupations of members of the first parliamentary term (1949-09-07 to 1953-09-07).

```
sort(table(members$beruf[members$wp == 1]), decreasing = T)[1:10]
```

```
## 
##               Landwirt           Rechtsanwalt               Kaufmann 
##                     20                     20                     15 
##           Angestellter Rechtsanwalt und Notar              Redakteur 
##                     13                     10                     10 
##   Bundesminister a. D.              Fabrikant               Hausfrau 
##                      9                      9                      8
##             Journalist
##                      8
```

And the code below returns the ten most common occupations of members of the current parliamentary term (since 2017-10-24).

```
sort(table(members$beruf[members$wp == 19]), decreasing = T)[1:10]
```

```
## 
##           Rechtsanwalt         Rechtsanwältin                 Jurist
##                     38                     14                     11
##               Juristin Politikwissenschaftler        Geschäftsführer
##                      9                      8                      7
##             Volljurist           Angestellter            Unternehmer
##                      7                      6                      6
##    Bürgermeister a. D.
##                      5
```

We see that occupations like farmer ("Landwirt") and housewife ("Hausfrau") have been replaced by new ones like political scientist ("Politikwissenschaftler"). Jurists were and continue to be very well represented in the Bundestag.

3.2 Teaching

btmembers can also be used by teachers. In data analysis classes, the package is well-suited for exercises on data visualization, clustered data, and textual analysis (especially with the `vita_kurz` variable, which is available for the last parliamentary term). It is also a great resource for classes on political representation and German politics. It provides a structured and easily accessible repertoire of facts about German parliamentary history.

4 Moving forward: Remaining problems and potential solutions

There are a few remaining problems with btmembers. I note here some of these limitations and suggest improvements for future versions of the package.

4.1 Speed

btmembers is slow. Running the function `import_members()` to update the data can take several minutes. I am currently in the process of integrating new functions from the package `dplyr` that should substantially increase the speed of the function.

4.2 Lost data

As explained previously, all the original data provided by the Bundestag is preserved, except previous names of members and their functions in the Bundestag. One way to keep all the data would be to give users the option to import not one but *multiple* dataframes, for example, one dataframe for names, one for biographical data, and one for functions. The dataframes could then be merged using the id and parliamentary term variables as merging keys. I am planning to implement this option in the near future.

4.3 Data update

The data that comes preloaded with btmembers is not updated automatically. Users have the option to run `update_available()` in combination with `import_members()` to get the latest version of the data from the Bundestag website. My long-term objective is to incorporate automatic checks on GitHub.

4.4 Problems with the original data

Some problems with btmembers do not relate to the package itself, but to the original data provided by the Bundestag. These problems are more difficult to fix.

1. **Some variables should have been coded as factors**. The data provided by the Bundestag was not intended to be condensed in a tabular dataset. Sometimes different values point to the same underlying concept. Family status, for example, has 60 different values: this can certainly be simplified. For the moment, I have refrained from recoding these variables as I am afraid I would loose some valuable information. Most variables have therefore been left as character instead of factor variables.

2. **The variable `partei_kurz` does not take into account changes in party affiliation**. It seems like the Bundestag only refers to the last affiliation, but this remains unclear.

3. **The variable `geburtsland` is not coded systematically**. In the original XML file, this variable was usually left empty when the member was born in Germany. Yet, the problem is that the borders of Germany changed over the course of the twentieth century and this is not reflected in the data. Should members born in Pomerania or Sudetenland be considered born in Germany? Also, the coding of countries of origin should follow an internationally agreed standard, such as ISO-3C. Creating an alternative variable might be necessary.

These limitations should be taken into consideration when using btmembers.

5 Conclusion

The Bundestag is arguably the most important democratic institution in Germany. Yet, with currently more than 700 members, examining the composition of this representative body can be challenging. The btmembers R package unpacks data on members of the Bundestag since 1949 and condenses it into a tabular dataset, a format better well-suited for comparative analyses. With federal elections scheduled in a few months, the package will help researchers, journalists, teachers, and the broader public get a clearer picture of the German parliament.