Description of Replication Material, Datasets, and R Scripts

Advocacy campaigns and gender bias in media coverage of elections

Fabrizio Gilardi Theresa Gessler Maël Kubli

In this file, we describe all datasets and R scripts required to rerun and reproduce the analyses reported in the paper and Supporting Information. If you have any questions, please do not hesitate to contact the authors.

Datasets

The repository for replicating the empirical analyses contains the following datasets. The following files (in alphabetic order) are used in the R scripts listed below. All other files in the repository, which are not listed here, are intermediate products of the analysis and are there to save time if only some steps of the analysis should be reproduced.

- *smd_ner_2015_2019_combined.RDS*: Main Dataset
- *diff_in_dff_m_vs_f_no_split_no_presidents.csv*: one dataset for diff. in diff. models
- *diff_in_diff_m_vs_f_no_split_presidents.csv:* one dataset for diff. in diff. models
- *diff_in_diff_m_vs_f_topic_split_no_presidents.csv:* one dataset for diff. in diff. models
- *diff_in_diff_m_vs_f_topic_split_presidents.csv:* one dataset for diff. in diff. models
- *diff_in_diff_m_vs_f_vs_NA_no_split_absolute_no_presidents.csv:* one dataset for diff. in diff. models
- *diff_in_diff_m_vs_f_vs_NA_no_split_absolute_presidents.csv:* one dataset for diff. in diff. models
- *diff_in_diff_m_vs_f_vs_NA_topic_split_no_presidents.csv:* one dataset for diff. in diff. models
- *diff_in_diff_m_vs_f_vs_NA_topic_split_presidents.csv:* one dataset for diff. in diff. models
- *big_newspapers_only_diff_in_diff_m_vs_f_no_split_no_presidents.csv:* one dataset for diff. in diff. appendix models
- *big_newspapers_only_diff_in_diff_m_vs_f_no_split_presidents.csv:* one dataset for diff. in diff. appendix models
- *big_newspapers_only_diff_in_diff_m_vs_f_topic_split_no_presidents.csv:* one dataset for diff. in diff. appendix models
- *big_newspapers_only_diff_in_diff_m_vs_f_topic_split_presidents.csv:* one dataset for diff. in diff. appendix models
- *big_newspapers_only_diff_in_diff_m_vs_f_vs_NA_no_split_absolute_no_presidents.csv:* one dataset for diff. in diff. appendix models
- *big_newspapers_only_diff_in_diff_m_vs_f_vs_NA_no_split_absolute_presidents.csv:* one dataset for diff. in diff. appendix models
- *big_newspapers_only_diff_in_diff_m_vs_f_vs_NA_topic_split_no_presidents.csv:* one dataset for diff. in diff. appendix models
- *big_newspapers_only_diff_in_diff_m_vs_f_vs_NA_topic_split_presidents.csv:* one dataset for diff. in diff. appendix models
- *PerClassEnsemblePerformance_FR.csv: Data for a table in appendix*
- *PerClassEnsemblePerformance.csv: Data for a table in appendix*

R Scripts

The repository contains the following R scripts. If you want to reproduce the entire analysis, please load and execute the scripts in the following order:

Classification Scripts (src_all_else/):

- *00-01-install_h2o_3_24_0_4.R*: script to install the h2o package with the right version number, with which all the models have been trained. To run this script you will need the following file:
    - *h2o_3.24.0.4.tar.gz*

classification_2019/ & classification_2015/

- *01-06-01-smd_data_classification.R*: scripts for 2019 and 2015 to classify all articles from newspapers in German. These scripts cannot run since the input data set is missing from the replication files. We cannot provide the newspaper articles for replication due to a binding data protection agreement with the Swiss media database. To run this script, if you have the Newspaper Articles Dataset, you will need the following files:
    - *W2V_2019_SMD_Model_600_DE*
    - *dl_01_de*
    - *dl_02_de*
    - *gbm_01_de*
    - *named_entities_swiss_politics_de.RDS*
    - *political_words_swiss_politics_de.RDS*
    - *non_political_words_swiss_politics_de.RDS*
    - *smd_all_curated.RDS* (Missing Dataset)
    - *SMD_TO_CLASSIFY.RDS* (Missing 2019 Dataset)
- *01-06-02-smd_data_classification.R*: scripts for 2019 and 2015 to classify all articles from newspapers in German. These scripts cannot run since the input data set is missing from the replication files. We cannot provide the newspaper articles for replication due to a binding data protection agreement with the Swiss media database. To run this script, if you have the Newspaper Articles Dataset, you will need the following files:
    - *W2V_2019_SMD_Model_600_FR*
    - *dl_01_fr*
    - *dl_02_fr*
    - *gbm_01_fr*
    - *named_entities_swiss_politics_fr.RDS*
    - *political_words_swiss_politics_fr.RDS*
    - *non_political_words_swiss_politics_fr.RDS*
    - *smd_all_curated.RDS* (Missing 2015 Dataset)
    - *SMD_TO_CLASSIFY.RDS* (Missing 2019 Dataset)
- *01-06-03-smd-data-classification.R*: scripts to merge the classified newspaper articles from 2019 and 2015 in German and French with each other. To run this script, you will need to run the smd_data_classification.R scripts from before first and have the files:
    - *smd_all_curated.rds* (2015)
    - *smd_to_classify.rds* (2019)
    - *smd_classified_other.rds* (2015 & 2019*)*
    - *smd_classified_fr.rds* (2015 & 2019)
    - *smd_classified_de.rds* (2015 & 2019)

sentiment_2015/ & sentiment_2019/

- 03-01-01-smd-data-sentiment-calculator.R: scripts to calculate sentiment for all articles for the 2019 and 2015 corpus. These scripts cannot run since the input data set is missing from the replication files. We cannot provide the newspaper articles for replication due to a binding data protection agreement with the Swiss media database. To run this script, if you have the Newspaper Articles Dataset, you will need the following files:
    - auto_dictionaries_lsd.RData
    - lsde_frence_gemane.RData
    - lsde_frenche_germane.RData
    - smd_all_curated_classified.RDS (Missing Dataset 2015 and 2019)

ner_2015/ & ner_2019/

- *04-01-01-smd-data-ner-preprocessing.R*: scripts to preprocess entity list for named entity recognition for 2015 and 2019. This script is executable and needs the following datasets:
    - *candidates_list_final_2019.csv*
    - *candidates_list_final_2015.csv*
- *04-02-01-smd-data-ner.R*: scripts to find and mark all named entities in the newspaper articles for 2019 and 2015. These scripts cannot run since the input data set is missing from the replication files. We cannot provide the newspaper articles for replication due to a binding data protection agreement with the Swiss media database. To run this script, if you have the Newspaper Articles Dataset, you will need the following files:
    - *smd_all_curated_classified_sentiment.RDS* (2015 & 2019)
    - *Named_Entity_List_withID* (2015 & 2019)
- *04-03-01-smd-ner-postprocessing.R*: script to postprocess named entity recognition data. To run this script, if you have the Newspaper Articles Dataset, you will need the following files:
    - *01-minified-corpus-NER_kandidaten_SMD.RDS* (2015 & 2019)
    - *00-Named_Entity_List_withID.csv* (2015 & 2019)
    - *00-partyList.RDS* (2015 & 2019)
    - *curated_party_regex.txt* (2015 & 2019)
- *04-04-01-smd-data-ner-merge.R*: script to merge named entity recognition data with corpus. These scripts cannot run since the input data set is missing from the replication files. We cannot provide the newspaper articles for replication due to a binding data protection agreement with the Swiss media database. To run this script, if you have the Newspaper Articles Dataset, you will need the following files:
- *02-minified-corpus_CURATED.RDS* (2015 & 2019)
- *smd_all_curated_classified_sentiment.RDS* 2015 & 2019)

Analysis Scrips (src_main/):

- *M0-01-Data-Preparation*: This script generates the combined dataset from 2015 and 2019 for the analysis. These scripts cannot run since the input data set is missing from the replication files. We cannot provide the newspaper articles for replication due to a binding data protection agreement with the Swiss media database. To run this script, if you have the Newspaper Articles Dataset, you will need the following files:
    - *2015/smd_all_curated_classified_sentiment.RDS* (Not Available)
    - *2019/smd_all_curated_classified_sentiment.RDS* (Not Available)
    - *2015/ner-SMD-annotated-corpus.RDS* (Not Available)
    - *2019/ner-SMD-annotated-corpus.RDS* (Not Available)

- *M0-02-Attention_Analysis_Loop.R*: This script contains the code to produce the diff in diff data in all its configurations with all newspapers. To run this script, you need the following files:
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*
- *M0-03-Attention_Analysis_Loop_Big_Newspapers_Only.R*: This script contains the code to produce the diff in diff data in all its configurations with only the "national" newspapers. To run this script, you need the following files:
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*
- *M1-01-Models.R*: This script runs the diff in diff models. To run this script, you need the following files:
    - *diff_in_diff_m_vs_f_no_split_no_presidents.csv*
    - *diff_in_diff_m_vs_f_vs_NA_no_split_absolute_no_presidents.csv*
    - *diff_in_diff_m_vs_f_topic_split_no_presidents.csv*
    - *diff_in_diff_m_vs_f_vs_NA_topic_split_no_presidents.csv*
- *M1-01Models-2.R*: This script runs the regression models. To run this scrip, you need the following files:
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*
- *M1-02-Figure_1.R*: This script generates the first plot of the article. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*
- *M1-03-Figure_2.R*: This script generates the second plot of the article. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*
- *M1-04-Figure_3.R*: This script generates the third plot of the article. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*

Appendix Scripts (src_appendix):

- *A1-01_ClassifierEvalTable.R*: This script generates the table with the Gilardi et al. 2022 classifier metrics. To run this script, you need the following files:
    - *PerClassEnsemblePerformance.csv*
    - *PerClassEnsemblePerformance_fr.csv*
- *A1-02_NewspaperStatslTable.R*: This script generates the newspaper statistics table in the appendix. To run this script, if you have the Newspaper Articles Dataset, you will need the following files:

    - *smd_ner_2015_2019_combined*
- *A1-02_TimelineElectionYear.R*: This script contains the code for figure afigure_0_1a.png. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*
- *A1-03_Average_Mentons.R*: This script contains the code for figure afigure_1_1a.png. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*
- *A1-04_Top_Mentioned_Female.R*: This script contains the code for figure afigure_2_1a.png. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*
- *A1-05_Female_Mentions_Party.R*: This script contains the code for figure afigure_3_1a.png. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*
- *A1-06_Dumbell.R*: This script contains the code for figure afigure_4_1a.png. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*
- *A1-07_Dumbell.R*: This script contains the code for figure afigure_5_1a.png. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*
- *A1-08_Dumbell.R*: This script contains the code for figure afigure_6_1a.png. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*
- *A1-09_Trends_mentions.R*: This script contains the code for figure afigure_7_1a.png. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*

- *A1-10_Trends_mentions_topics.R*: This script contains the code for figure afigure_8_1a.png. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*
- *A1-11_Female_Candidates_Party.R*: This script contains the code for figure afigure_9_1a.png and afigure_9_2a.png. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*
    - */support/parliament/frauenanteil_nr_zeitreihe_2.csv*
- *A1-12_Female_Candidates_over_Time_NR.R*: This script contains the code for figure afigure_10_1a.png. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *frauenanteil_nr_zeitreihe.csv*
- *A1-13_Party_Strength_Over_Time.R*: This script contains the code for figure afigure_11_1a.png. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *frauenanteil_nr_zeitreihe_2.csv*
- *A1-14_Party_Strength_Cantons.R*: This script contains the code for figure afigure_12_1a.png. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *frauenanteil_nr_zeitreihe_2.csv*
- *A1-15_hits_newspaper_descriptive.R*: This script contains the code for figure afigure_13_1a.png and afigure_13_2a.png. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*
- *A1-16_models_2019_only_1.R*: This script contains the code for the model tables showing just results for 2019. To run this script, you need the following files:
    - *diff_in_diff_m_vs_f_no_split_no_presidents.csv*
    - *diff_in_diff_m_vs_f_vs_NA_no_split_absolute_no_presidents.csv*
    - *diff_in_diff_m_vs_f_topic_split_no_presidents.csv*
    - *diff_in_diff_m_vs_f_vs_NA_topic_split_no_presidents.csv*
- *A1-17_models_2019_only_2.R*: This script contains the code for the model tables showing just results for 2019. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*
- *A1-18_models_national_newspapers_only_1.R*: This script contains the code for the model tables showing only results with big newspapers. To run this script, you need the following files:
    - *diff_in_diff_m_vs_f_no_split_no_presidents.csv*
    - *diff_in_diff_m_vs_f_vs_NA_no_split_absolute_no_presidents.csv*

    - *diff_in_diff_m_vs_f_topic_split_no_presidents.csv*
    - *diff_in_diff_m_vs_f_vs_NA_topic_split_no_presidents.csv*
- *A1-19_models_national_newspapers_only_2.R*: This script contains the code for the model tables showing only results with big newspapers. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - *smd_ner_2015_2019_combined.RDS*
    - */support/candidates-2015/00-Named_Entity_List_withID.csv*
    - */support/candidates-2019/00-Named_Entity_List_withID.csv*
- *A1-20_less_female_incumbents_sr_2019.R*: This script contains the code producing several numbers shown in the paper. To run this script, you need the following files:
    - *ggplot_theme_ddl.R*
    - */support/candidates-2019/2019_chvote_councilofstates.csv*
    - */support/candidates-2015/candidates_list_final_2015.csv*
    - */support/candidates-2019/candidates_list_final_2019.csv*

The beginning of each R script lists the packages that need to be loaded and the versions used when running the scripts.
The entire analysis has been successfully executed using the following platform and package versions:

```
R version 4.4.1 (2024-06-14)
Platform: x86_64-apple-darwin20
Running under: macOS Sonoma 14.5

Matrix products: default
BLAS:
/System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framew
ork/Versions/A/libBLAS.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/4.4-
x86_64/Resources/lib/libRlapack.dylib;  LAPACK version 3.12.0

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

time zone: Europe/Zurich
tzcode source: internal

attached base packages:
[1] grid      stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] forcats_1.0.0   purrr_1.0.2     tidyr_1.3.1     tibble_3.2.1
tidyverse_2.0.0
 [6] pacman_0.5.1    stringr_1.5.1   ggalt_0.4.0     MASS_7.3-60.2   texreg_1.39.3
[11] effects_4.2-2   carData_3.0-5   lme4_1.1-35.5   Matrix_1.7-0    sjPlot_2.8.16
[16] ggpubr_0.6.0    gridExtra_2.3   readxl_1.4.3    png_0.1-8       ggforce_0.4.2
[21] here_1.0.1      reshape2_1.4.4  extrafont_0.19  lubridate_1.9.3 scales_1.3.0
[26] ggrepel_0.9.5   ggridges_0.5.6  ggExtra_0.10.1  ggeffects_1.7.0
gganimate_1.0.9
[31] ggplot2_3.5.1   dplyr_1.1.4     readr_2.1.5

loaded via a namespace (and not attached):
 [1] DBI_1.2.3          rlang_1.1.4        magrittr_2.0.3     compiler_4.4.1
 [5] vctrs_0.6.5        maps_3.4.2         pkgconfig_2.0.3    crayon_1.5.3
 [9] fastmap_1.2.0      backports_1.5.0    utf8_1.2.4         promises_1.3.0
[13] tzdb_0.4.0         nloptr_2.1.1       xfun_0.45          ash_1.0-15
[17] progress_1.2.3     later_1.3.2        sjmisc_2.8.10      tweenr_2.0.3
[21] broom_1.0.6        prettyunits_1.2.0  R6_2.5.1           RColorBrewer_1.1-3
[25] stringi_1.8.4      car_3.1-2          boot_1.3-30        extrafontdb_1.0
[29] cellranger_1.1.0   Rcpp_1.0.12        knitr_1.47         httpuv_1.6.15
```

```
[33] splines_4.4.1       nnet_7.3-19         timechange_0.3.0    tidyselect_1.2.1
[37] rstudioapi_0.16.0   abind_1.4-5         miniUI_0.1.1.1      sjlabelled_1.2.0
[41] lattice_0.22-6      plyr_1.8.9          shiny_1.8.1.1       withr_3.0.0
[45] survival_3.6-4      survey_4.4-2        polyclip_1.10-6     pillar_1.9.0
[49] KernSmooth_2.23-24 insight_0.20.1      generics_0.1.3      rprojroot_2.0.4
[53] hms_1.1.3           munsell_0.5.1       minqa_1.2.7         xtable_1.8-4
[57] glue_1.7.0          tools_4.4.1         ggsignif_0.6.4      cowplot_1.1.3
[61] mitools_2.4         Rttf2pt1_1.3.12     datawizard_0.11.0   colorspace_2.1-0
[65] nlme_3.1-164        performance_0.12.0 cli_3.6.3           proj4_1.0-14
[69] fansi_1.0.6         sjstats_0.19.0      gtable_0.3.5        rstatix_0.7.2
[73] digest_0.6.36       farver_2.1.2        htmltools_0.5.8.1   lifecycle_1.0.4
[77] httr_1.4.7          mime_0.12
```