Lecture_4_Notes

Athena Zhang

2023-09-28

Data Wrangling

0. Load the tidyverse package

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

1. Import the V-Dem Data

```
d <- read_csv("_DataPublic_/vdem/1984_2022/vdem_1984_2022_external.csv")
```

```
## Rows: 6789 Columns: 211
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr    (3): country_name, country_text_id, histname
## dbl  (207): country_id, year, project, historical, codingstart, codingend, c...
## date   (1): historical_date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
> d
# A tibble: 6,789 × 211
   country_name country_text_id country_id  year historical_date project historical histname            codingstart codingend
   <chr>        <chr>                <dbl> <dbl> <date>            <dbl>      <dbl> <chr>                     <dbl>     <dbl>
 1 Mexico       MEX                      3  1984 1984-12-31            0          1 United Mexican Sta…        1789      2022
 2 Mexico       MEX                      3  1985 1985-12-31            0          1 United Mexican Sta…        1789      2022
 3 Mexico       MEX                      3  1986 1986-12-31            0          1 United Mexican Sta…        1789      2022
 4 Mexico       MEX                      3  1987 1987-12-31            0          1 United Mexican Sta…        1789      2022
 5 Mexico       MEX                      3  1988 1988-12-31            0          1 United Mexican Sta…        1789      2022
 6 Mexico       MEX                      3  1989 1989-12-31            0          1 United Mexican Sta…        1789      2022
 7 Mexico       MEX                      3  1990 1990-12-31            0          1 United Mexican Sta…        1789      2022
 8 Mexico       MEX                      3  1991 1991-12-31            0          1 United Mexican Sta…        1789      2022
 9 Mexico       MEX                      3  1992 1992-12-31            0          1 United Mexican Sta…        1789      2022
10 Mexico       MEX                      3  1993 1993-12-31            0          1 United Mexican Sta…        1789      2022
# ℹ 6,779 more rows
# ℹ 201 more variables: codingstart_contemp <dbl>, codingend_contemp <dbl>, codingstart_hist <dbl>, codingend_hist <dbl>,
#   gapstart1 <dbl>, gapstart2 <dbl>, gapstart3 <dbl>, gapend1 <dbl>, gapend2 <dbl>, gapend3 <dbl>, gap_index <dbl>,
#   COWcode <dbl>, e_v2x_api_3C <dbl>, e_v2x_api_4C <dbl>, e_v2x_api_5C <dbl>, e_v2x_civlib_3C <dbl>, e_v2x_civlib_4C <dbl>,
#   e_v2x_civlib_5C <dbl>, e_v2x_clphy_3C <dbl>, e_v2x_clphy_4C <dbl>, e_v2x_clphy_5C <dbl>, e_v2x_clpol_3C <dbl>,
#   e_v2x_clpol_4C <dbl>, e_v2x_clpol_5C <dbl>, e_v2x_clpriv_3C <dbl>, e_v2x_clpriv_4C <dbl>, e_v2x_clpriv_5C <dbl>,
#   e_v2x_corr_3C <dbl>, e_v2x_corr_4C <dbl>, e_v2x_corr_5C <dbl>, e_v2x_cspart_3C <dbl>, e_v2x_cspart_4C <dbl>, …
# ℹ Use `print(n = ...)` to see more rows, and `colnames()` to see all variable names
```