

Extra Credit

Heleine Fouda

2023-11-05

ASSIGNMENT: Working with the two JSON files available through the API at nobelprize.org, ask and answer 4 interesting questions, e.g. “Which country”lost” the most nobel laureates (who were born there but received their Nobel prize as a citizen of a different country)?”

Getting started

Retrieve data from the API and parse it as JSON:

Various options for importing the Json file:

```
# Option 1
api_url <- "https://api.nobelprize.org/2.1/laureates?_ga=2./1901-2023"

response <- GET(api_url)
response
```

```
## Response [https://api.nobelprize.org/2.1/laureates?_ga=2./1901-2023]
##   Date: 2023-11-06 02:38
##   Status: 200
##   Content-Type: application/json;charset=utf-8
##   Size: 90.7 kB
```

```
# option 2
api_url1 <- "https://api.nobelprize.org/2.1/laureates?_ga=2.155154616.1232644105.1699062535-556212721.1699062535-556212721"
api_url2 <- "https://api.nobelprize.org/2.1/nobelPrizes?_ga=2.123220747.1232644105.1699062535-556212721.1699062535-556212721"
full_url <- paste0(api_url1,api_url2)
response <- GET(full_url)

response
```

```
## Response [https://api.nobelprize.org/2.1/laureates?_ga=2.155154616.1232644105.1699062535-556212721.1699062535-556212721]
##   Date: 2023-11-06 02:38
##   Status: 200
##   Content-Type: application/json;charset=utf-8
##   Size: 90.7 kB
```

```
# Option 3, retrieving the data set from the github repository
prize_url <- "https://raw.githubusercontent.com/Heleinef/Data-Science-Master_Heleine/main/Laureates"
prize_json <- GET(prize_url)

prize_json
```

```
## Response [https://raw.githubusercontent.com/Heleinef/Data-Science-Master_Heleine/main/Laureates]
##   Date: 2023-11-06 02:38
##   Status: 200
```

```
## Content-Type: text/plain; charset=utf-8
## Size: 119 kB
## {"laureates":[{"id":"745","knownName":{"en":"A. Michael Spence","se":"A. Mich...
## {"nobelPrizes":[{"awardYear":"1901","category":{"en":"Chemistry","no":"Kjemi"...

# Check if the request was successful
prize_json$status_code

## [1] 200
```

Converting the JSON file into a data frame:

```
nobel_data <- content(response, "text", encoding = "UTF-8")
nobel_df <- jsonlite::fromJSON(nobel_data)
```

```
# Summary
summary(nobel_df)
```

```
##           Length Class      Mode
## laureates 14      data.frame list
## meta       6      -none-      list
## links      4      -none-      list
```

```
Nobels_df <- as.data.frame(nobel_df)
tibble(Nobels_df)
```

```
## # A tibble: 25 x 24
##   laureates.id laureates.knownName$en $se laureates.givenName$en
##   <chr>         <chr>                <chr>         <chr>
## 1 745          A. Michael Spence      A. Michael Spence A. Michael
## 2 102          Aage N. Bohr           Aage N. Bohr     Aage N.
## 3 779          Aaron Ciechanover      Aaron Ciechanover Aaron
## 4 259          Aaron Klug              Aaron Klug        Aaron
## 5 1004         Abdulrazak Gurnah      Abdulrazak Gurnah Abdulrazak
## 6 114          Abdus Salam            Abdus Salam       Abdus
## 7 982          Abhijit Banerjee       Abhijit Banerjee Abhijit
## 8 981          Abiy Ahmed Ali         Abiy Ahmed Ali    Abiy
## 9 843          Ada E. Yonath          Ada E. Yonath     Ada E.
## 10 866         Adam G. Riess          Adam G. Riess     Adam G.
## # i 15 more rows
## # i 22 more variables: laureates.givenName$se <chr>,
## # laureates.familyName <df[,2]>, laureates.fullName <df[,2]>,
## # laureates.fileName <chr>, laureates.gender <chr>, laureates.birth <df[,2]>,
## # laureates.wikipedia <df[,2]>, laureates.wikidata <df[,2]>,
## # laureates.sameAs <list>, laureates.links <list>,
## # laureates.nobelPrizes <list>, laureates.death <df[,2]>, ...
```

Our working data set from the retrieved data

```
Nobels_laureates <- Nobels_df %>%
  select(laureates.id, laureates.fileName, laureates.gender, links.first)
Nobels_laureates
```

```
##   laureates.id laureates.fileName laureates.gender
## 1           745           spence           male
## 2           102           bohr            male
## 3           779       ciechanover           male
```

## 4	259	klug	male
## 5	1004	gurnah	male
## 6	114	salam	male
## 7	982	banerjee	male
## 8	981	abiy	male
## 9	843	yonath	female
## 10	866	riess	male
## 11	199	butenandt	male
## 12	164	baeyer	male
## 13	185	windaus	male
## 14	541	esquivel	male
## 15	292	zewail	male
## 16	853	suzuki	male
## 17	978	yoshino	male
## 18	819	gore	male
## 19	1012	aspect	male
## 20	729	heeger	male
## 21	376	hodgkin	male
## 22	730	macdiarmid	male
## 23	11	michelson	male
## 24	628	camus	male
## 25	403	claudio	male

links.first

## 1	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 2	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 3	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 4	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 5	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 6	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 7	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 8	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 9	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 10	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 11	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 12	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 13	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 14	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 15	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 16	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 17	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 18	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 19	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 20	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 21	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 22	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 23	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 24	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25
## 25	https://masterdataapi.nobelprize.org/2.1/laureates?offset=0&limit=25

laureates names

```
Nobels_name <- Nobels_df %>%
  select(laureates.fileName)
Nobels_name
```

```
## laureates.fileName
## 1      spence
## 2      bohr
## 3      ciechanover
## 4      klug
## 5      gurnah
## 6      salam
## 7      banerjee
## 8      abiy
## 9      yonath
## 10     riess
## 11     butenandt
## 12     baeyer
## 13     windaus
## 14     esquivel
## 15     zewail
## 16     suzuki
## 17     yoshino
## 18     gore
## 19     aspect
## 20     heeger
## 21     hodgkin
## 22     macdiarmid
## 23     michelson
## 24     camus
## 25     claude
```

Laureates ids

```
Nobels_id<- Nobels_df %>%
  select(laureates.id)
Nobels_id
```

```
## laureates.id
## 1      745
## 2      102
## 3      779
## 4      259
## 5     1004
## 6      114
## 7      982
## 8      981
## 9      843
## 10     866
## 11     199
## 12     164
## 13     185
## 14     541
## 15     292
## 16     853
## 17     978
## 18     819
## 19    1012
## 20     729
## 21     376
```

```
## 22      730
## 23      11
## 24     628
## 25     403
```

Gender distribution

```
nobels_gender <- Nobels_df %>%
  select(laureates.gender)
nobels_gender
```

```
## laureates.gender
## 1      male
## 2      male
## 3      male
## 4      male
## 5      male
## 6      male
## 7      male
## 8      male
## 9      female
## 10     male
## 11     male
## 12     male
## 13     male
## 14     male
## 15     male
## 16     male
## 17     male
## 18     male
## 19     male
## 20     male
## 21     male
## 22     male
## 23     male
## 24     male
## 25     male
```

```
Nobels_laureates <- Nobels_df %>%
  select(laureates.id, laureates.fileName, laureates.gender)
ggplot(Nobels_laureates, aes(y = laureates.id, color = laureates.gender)) +
  geom_bar()
```

