Lab 03 - Nobel laureates

Your Name

Contents

In January 2017, Buzzfeed published an article on why Nobel laureates show immigration is so important for American science. You can read the article here. In the article they show that while most living Nobel laureates in the sciences are based in the US, many of them were born in other countries. This is one reason why scientific leaders say that immigration is vital for progress. In this lab we will work with the data from this article to recreate some of their visualizations as well as explore new questions.

1 Learning goals

- Replicating published results
- Data wrangling and visualisation

2 Lab prep

Read the Buzzfeed article titled *These Nobel Prize Winners Show Why Immigration Is So Important For American Science.* We will replicate this analysis in the workshop so it's crucial that you're familiar with it ahead of time.

3 Getting started

Go to the course GitHub organization and locate your lab repo, which should be named `lab-03-nobel-laureates-YOUR_GITH`
Grab the URL of the repo, and clone it in RStudio. First, open the R Markdown document `lab-03.Rmd` and Knit it. Make sure it compiles without errors. The output will be in the file markdown `.md` file with the same name.

3.1 Warm up

Before we introduce the data, let's warm up with some simple exercises.

- Update the YAML, changing the author name to your name, and **knit** the document.
- Commit your changes with a meaningful commit message.
- Push your changes to GitHub.
- Go to your repo on GitHub and confirm that your changes are visible in your Rmd **and** md files. If anything is missing, commit and push again.

3.2 Packages

We'll use the **tidyverse** package for much of the data wrangling. This package is already installed for you. You can load them by running the following in your Console:

```
library(tidyverse)
```

3.3 Data

The dataset for this assignment can be found as a CSV (comma separated values) file in the `data` folder of your repository. You can read it in using the following.

```
nobel <- read_csv("data/nobel.csv")
```

The variable descriptions are as follows:

- `id`: ID number
- `firstname`: First name of laureate
- `surname`: Surname
- `year`: Year prize won
- `category`: Category of prize
- `affiliation`: Affiliation of laureate
- `city`: City of laureate in prize year
- `country`: Country of laureate in prize year
- `born_date`: Birth date of laureate
- `died_date`: Death date of laureate
- `gender`: Gender of laureate
- `born_city`: City where laureate was born
- `born_country`: Country where laureate was born
- `born_country_code`: Code of country where laureate was born
- `died_city`: City where laureate died
- `died_country`: Country where laureate died
- `died_country_code`: Code of country where laureate died
- `overall_motivation`: Overall motivation for recognition
- `share`: Number of other winners award is shared with
- `motivation`: Motivation for recognition

In a few cases the name of the city/country changed after laureate was given (e.g. in 1975 Bosnia and Herzegovina was called the Socialist Federative Republic of Yugoslavia). In these cases the variables below reflect a different name than their counterparts without the suffix '_original'.

- `born_country_original`: Original country where laureate was born
- `born_city_original`: Original city where laureate was born
- `died_country_original`: Original country where laureate died
- `died_city_original`: Original city where laureate died
- `city_original`: Original city where laureate lived at the time of winning the award
- `country_original`: Original country where laureate lived at the time of winning the award