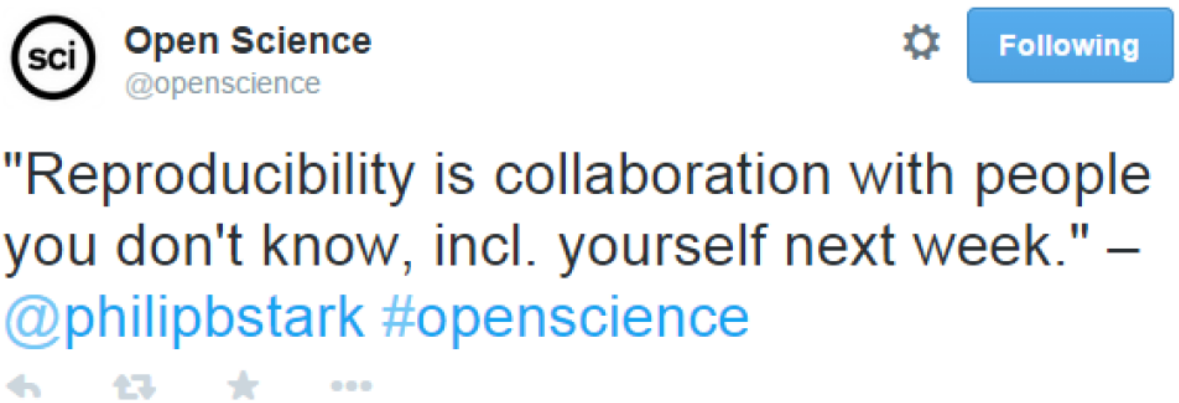
A Populist Crisis? Lab Session

Julia Schulte-Cloos
European University Institute
julia.schulte-cloos@eui.eu
jschultecloos.github.io

12 July 2018

Why Reproducible Research?

- Benefits yourself! 'Create a better relationship with your future self' (Bowers and Voors 2016)



- That's where social science is moving to!
  - Journals and funding agency requirements
  - Open science and data sharing
  - Confidence in your own work, easier colloboration and workflow

A definition of reproducible research?

What is **irr**reproducible research?

- Bad data, e.g. entered or manually edit by humans.
  - Consider the dog licensing database for Cook County, Illinois
- 'Available from the author'
- PloS: software created for use in publications must be open source
- **Reproducibility vs. replication**
- **Literate programming** & communication via code

A definition of reproducible research?

What is **irr**reproducible research?

- Bad data, e.g. entered or manually edit by humans.
  - Consider the dog licensing database for Cook County, Illinois
  - Contains at least 250 spellings of Chihuahua
- 'Available from the author'
- PloS: software created for use in publications must be open source
- **Reproducibility vs. replication**
- **Literate programming** & communication via code

Getting started

- **Markdown** as a human readable way to style text
- **Pandoc**:
  - allows for conversion between different markup languages
  - converts between docx, HTML, LaTeX, and Markdown
- Output: we focus on Knitr and LaTeX. Not only for reasons of its beauty.

The Setup

The YAML header

- sets global parameters of the document
- `title, author, date, fontsize, geometry, linestretch, abstract`
- customization according to LaTeX preferences:
  - packages can be loaded in YAML header after `header-includes:`
  - Pandoc variables for LaTeX

The Setup

The YAML header

- sets global parameters of the document
- `title, author, date, fontsize, geometry, linestretch, abstract`
- customization according to LaTeX preferences:
  - packages can be loaded in YAML header after `header-includes:`
  - Pandoc variables for LaTeX
- **Cheatsheet** for RMarkdown and official reference guide

Formatting

- **Text Formatting and Emphasis**
  - Bold text can be created with `**bold text**` or equivalently `__bold__`
  - Italic text can be created with `*italic*` or equivalently `_italic_`
- **Sections, Lists**

```
# A level-one section
## A level-two section with a [link](/url)
# An unnumbered section {-}, or equivalently # An unnumbered section
{.unnumbered}
```

- Always include blank line before a header
- Sections can be labelled and referenced by including an attribute after the header `{#sec:introduction}`

Footnotes and Images

- **Footnotes**
  - include inline footnotes with `^[footnote]`
  - multi-paragraph footnotes
- **Embedding images**
  - `![A nice elephant](./figs/elephant.jpg){width=90%}`
  - or: `\includegraphics[width=0.9\textwidth]{./figs/elephant.jpg}`
  - set global options for the inclusion of figures in the YAML header

```
---
output:
  pdf_document:
fig_width: 7
fig_height: 6
---
```

Beamer Presentations

- **Themes**
  - see also this and this gallery.

```
output:
  beamer_presentation:
    theme: "AnnArbor"
    colortheme: "dolphin"
    fonttheme: "structurebold"
    includes:
      in_header: beamer-header.txt
```

The setup and loading packages into R

The European Social Survey

We have downloaded the latest round of the ESS data, round 8 (released on 31/10/2017). We will first explore the data. The data comes only in SPSS format or STATA format. Unfortunately, both statistical softwares are not open source. Fortunately, you can use R to import the data even without being in posession of a licence for those two software packages.

Download data

Fantastic work by some data scientists, who wrote a package in R that downloads the ESS data automatically. See the website here.

```
# library(essurvey)
# show_countries()
# set_email("your.ess.registred.email.adress@email.com")
# eight_round =
#  import_rounds(8) %>%
#  recode_missings()
# write.csv(eight_round, file="ess8.csv", fileEncoding = "UTF8")
```

Import Data

```
ess_data <- read_dta("ESS8e02.dta")

ess_data = recode_missings(ess_data)
```

Navigating in the large dataset

How to navigate in a large dataset with 536 columns, i.e. variables?

```
# find_var(ess_data, pattern="country")
# find_var(ess_data, pattern="gender")
# find_var(ess_data, pattern="immigrant")
```

Explorative analysis

We first explore how many respondents are included in the dataset for each of the countries. If we want to compare across countries, we have to apply weights to the data. Read more about the weights in the ESS here.
How many countries are included in our dataset?

```
kable(ess_data %>%
        group_by(cntry) %>%
        summarise(observations= n()),
      caption = "Number of observations per country", format = "latex", b
  kable_styling(font_size = 9)
```

Table 1: Number of observations per country

cntry	observations
AT	2010
BE	1766
CH	1525
CZ	2269
DE	2852
EE	2019

Cross-Country Analysis

```
# find_var(ess_data, pattern="immigrant")
# table(ess_data$imwbcnt)
kable(ess_data %>%
        dplyr::filter(imwbcnt==0) %>%
        group_by(cntry) %>%
        summarise(N_immigrationworse = n()), format = "latex", booktabs =
      caption = "N of respondents thinking immigrants make country worse
  kable_styling(font_size = 9)
```

Table 2: N of respondents thinking immigrants make country worse place to live

cntry	N_immigrationworse
AT	204
BE	57
CH	17
CZ	264
DE	115
EE	139
ES	56

Immgriation Scepticism

Next, we can look at the share of people who believe that immigration is making the country a worse place to live. We first create an "indicator variable", that takes the value 1 if the respondent thinks immigration makes the country a totally worse place to live (value 0), and takes the value 0 if the respondent does not hold this extreme negative position. Variables that only take either the value 0 or the value 1 are also called binary variables, or dummy variables.

Immgriation Scepticism

We first create an "indicator variable", that takes the value 1 if the respondent thinks immigration makes the country a totally worse place to live (value 0), and takes the value 0 if the respondent does not hold this extreme negative position.

```
# this is a so-called "if_else" condition.
ess_data$immigrationworse = if_else(ess_data$imwbcnt==0, 1, 0)

kable(ess_data  %>%
      group_by(cntry) %>% filter(!is.na(immigrationworse)) %>%
      summarise(Share_immigrationworse = mean(immigrationworse)),
      caption = "Share of respondents thinking immigrants make country ar
      digits=2, format = "latex", booktabs = T) %>%
  kable_styling(font_size = 9)
```

Table 3: Share of respondents thinking immigrants make country an absolutely worse place to live

cntry	Share_immigrationworse
AT	0.10
BE	0.03
CH	0.01

Pro-Immigration Attitudes

We can also look at the share of respondents thinking that immigration make country a completely better place to live to contrast it. We again create an "indicator variable".

```
ess_data$immigrationbetter = if_else(ess_data$imwbcnt==10, 1, 0)
kable(ess_data  %>%
      group_by(cntry) %>% filter(!is.na(immigrationbetter)) %>%
      summarise(Share_immigrationbetter = mean(immigrationbetter)),
      caption = "Share of respondents thinking immigrants make country ar
      digits=2, format = "latex", booktabs = T) %>%
  kable_styling(font_size = 9)
```

Table 4: Share of respondents thinking immigrants make country an absolutely better place to live

cntry	Share_immigrationbetter
AT	0.02
BE	0.01
CH	0.03
CZ	0.01
DE	0.04

Average Attitude towards Immigration

Finally, we can also look at the average opinion of respondents on whether or not immigration makes their country a better place to live. The outcome variable is measured on a 0 to 10 scale, which we need to know if we want to interpret the different country-specific values.

Average Attitude towards Immigration

Finally, we can also look at the average opinion of respondents on whether or not immigration makes their country a better place to live.

```
kable(ess_data  %>%
      group_by(cntry) %>% filter(!is.na(imwbcnt)) %>%
      summarise(Avg_Immigrationworse = mean(imwbcnt)),
      caption = "Avg. attitude whether immigration makes country a better
      digits=2, format = "latex", booktabs = T) %>%
  kable_styling(font_size = 9)
```

Table 5: Avg. attitude whether immigration makes country a better / worse place to live (0-10 scale)

cntry	Avg_Immigrationworse
AT	4.19
BE	5.12
CH	5.48
CZ	3.68
DE	5.21
EE	4.26

Combining the different statistics

Instead of producing endlessly many tables that include only the respective mean for **one variable at a time**, we might consider to produce a table that includes all our values next to each other. This way, we can compare how high the share of people is that think immigration makes their country a completely worse place to live, a completely better place to live and the respective average value.

Combining the different statistics

```
kable(ess_data  %>%
      group_by(cntry) %>% filter(!is.na(immigrationbetter) & !is.na(immig
      select(cntry, immigrationbetter, immigrationworse, imwbcnt) %>%
      summarise_all(funs(mean), na.rm = TRUE),
      caption = "Share of respondents thinking immigrants make country be
      digits=2,
      col.names = c("Country", "Share Immigration Country Much Better", "
  kable_styling(font_size = 9)
```

Table 6: Share of respondents thinking immigrants make country better place to live

Country	Share Immigration Country Much Better	Share Immigration Country Muc
AT	0.02	
BE	0.01	
CH	0.03	
CZ	0.01	
DE	0.04	
EE	0.01	

Polarization of Anti-Immigration Attitudes

What we have been doing thus far substantively speaking, is that we looked at something that we might want to call the "polarization" of voters' attitude on the issue of immigration. We have considered the share of people who are convinced that immigration makes their country a **completely** worse place to live in, on the one hand, and we have considered the share of people who think that immigration makes their country a **completely** better place to live in. We have then contrasted these two values with the average opinion towards the issue within a country (measured on a 0-10 scale).

Country Analysis: Italy

Let us next look at one country and explore the characteristics of voters that are in favour of immigration or oppose it. We will focus on Italy, the country that had the highest share of respondents that believe that immigration makes their country a completely worse place to live.
Let us first create an indicator variable that measures "immigration sceptic" attitudes more generally, going beyond the extreme values of 10 and 0 that we used before. We will split the index in the middle. An immigration sceptic is someone who scores from 0 to 4 on his attitude on whether immigration is good for Italy and someone who is located at the upper end of the scale is not considered an anti-immigrant person.

Country Analysis: Italy

As we will do some more analyses only considering Italy, it may be useful to create a subset of the data that only contains Italy.

```
italy_data = ess_data %>% filter(cntry=="IT")
italy_data$antiimmigration = if_else(italy_data$imwbcnt<5, 1, 0)

kable(italy_data %>%
      filter(!is.na(antiimmigration)) %>%
      summarise(antiimmigration_share = mean(antiimmigration)), format =
  kable_styling(font_size = 9)
```

antiimmigration_share
0.5896524

Anti-Immigrant Attitudes and Socio-Demographic Characteristics

Let us first find some variables that could be interesting in terms of socio-economic characteristics.

```
# find_var(italy_data, pattern="age")
# find_var(italy_data, pattern="paid work")
# find_var(italy_data, pattern="vote")
# table(italy_data$agea)
# table(italy_data$pdwrk)
# find_var(italy_data, pattern="region")
# create North-South variable, see: https://en.wikipedia.org/wiki/NUTS_st
italy_data$south = if_else((italy_data$region=="ITF1" | italy_data$regio
                            italy_data$region=="ITF3" | italy_data$regio
                            italy_data$region=="ITF5" |  italy_data$regi
# table(italy_data$south)
```

Anti-Immigration and Socio-Demographics

```
kable(italy_data  %>%
        group_by(south, antiimmigration) %>%
        filter(!is.na(antiimmigration)) %>%
      select(agea, pdwrk, south, antiimmigration) %>%
      summarise_all(funs(mean), na.rm = TRUE),
      caption = "Socio-demographic characteristics of immigration-sceptic
      digits=2, format = "latex", booktabs = T,
      col.names = c("South", "Anti-Immigration 0/1", "Age", "Job 0/1")) %
  kable_styling(font_size = 9)
```

Table 7: Socio-demographic characteristics of immigration-sceptics in Italy

South	Anti-Immigration 0/1	Age	Job 0/1
0	0	45.90	0.54
0	1	52.44	0.49
1	0	44.20	0.42
1	1	48.28	0.36

Anti-Immigrant Attitudes and Political Behaviour

Finally, we would also like to know how the anti-immgrant attitudes among Italian voters relate to their voting behaviour, their willigness to stay in the EU, and their previous electoral participation. Therefore, we will focus on **eligible voters**.

Subsetting Eligible Voters

We subset the data to include only those who are 18 and were eligible to vote during the last election.

```
# find_var(italy_data, pattern="vote")
italy_eligible_data = italy_data %>% filter(agea>=18 & vote!=3)
# find_var(italy_eligible_data, pattern="Italy")
```

As the last national election at the time of the interview was a while ago (in 2013), let us rather consider the proximity to a certain party.

```
# table(italy_eligible_data$prtclcit) # Movimento 5stelle: 1, Lega Nord:
italy_eligible_data$lega = if_else(italy_eligible_data$prtclcit==3, 1, 0)
italy_eligible_data$movimento = if_else(italy_eligible_data$prtclcit==1,
# vote ==2 means that respondent did not vote
italy_eligible_data$vote = if_else(italy_eligible_data$vote==1, 1, 0)
# table(italy_eligible_data$vote)
```

Preferences to Leave the EU

Finally, let's consider whether a respondent would vote to leave the EU.

```
# table(italy_eligible_data$vteurmmb)
# 2 means voter would vote for leaving the EU
italy_eligible_data$leaveEU = if_else(italy_eligible_data$vteurmmb==2, 1,
# table(italy_eligible_data$leaveEU)
```

Anti-Immigration and Political Behaviour

```
kable(italy_eligible_data  %>%
        group_by(south, antiimmigration) %>%
        filter(!is.na(antiimmigration)) %>%
      select(agea, pdwrk, lega, movimento, leaveEU, south, antiimmigratio
      summarise_all(funs(mean), na.rm = TRUE),
      caption = "Socio-demographic characteristics and political attitude
      digits=2,
      col.names = c("South", "Anti-Immigr.", "Age", "Job", "Close to Lega
  kable_styling(font_size = 9)
```

Table 8: Socio-demographic characteristics and political attitudes among anti- and pro-immigration respondents in Italy

South	Anti-Immigr.	Age	Job	Close to Lega	Close to M5S	Leave EU
0	0	49.45	0.59	0.10	0.24	0.11
0	1	54.71	0.51	0.22	0.33	0.38
1	0	48.05	0.44	0.02	0.44	0.16
1	1	49.86	0.38	0.09	0.36	0.38

Summing Up: Analysis of ESS Round 8

These numbers are **basic descriptive** statistics. As we are working with observational data and cannot be confident to make claims about **causality**, these statistics give us, however, already a very good understanding of differences in socio-demographic characteristics and political attitudes among immigration sceptics and non-immigration sceptics in Italy.

Kosovo in the European Social Survey (Round 6)

We have downloaded round 6 of the ESS data that is the only round that includes data on Kosovo. Respondents were interviewed in February and March 2013.

```
ess_kosovo = read_dta("kosovo_ess.dta")

head(ess_kosovo)
```

Socio-Economic Characteristics of VV Supporters in 2013

```
# who feels close to VV?
# table(ess_kosovo$prtclxk) # vv is number 3
ess_kosovo$vv = if_else(ess_kosovo$prtclxk==3, 1, 0)

# find_var(ess_kosovo, pattern="age")
# find_var(ess_kosovo, pattern="education")
# table(ess_kosovo$edulvlb) # highest level of education
ess_kosovo$female = if_else(ess_kosovo$gndr==2, 1, 0)
ess_kosovo$higheducation = if_else(ess_kosovo$edulvlb> 323, 1, 0)
```

Socio-Economic Characteristics of VV Supporters in 2013

```
kable(ess_kosovo  %>%
      group_by(vv) %>%
      select(agea, female, higheducation, pdwrk, vv) %>%
      summarise_all(funs(mean), na.rm = TRUE),
      caption = "Average socio-economic characteristics of VV partisans",
      digits=2, format = "latex", booktabs = T,
      col.names = c("VV", "Age", "Female", "Higher Education", "Job")) %>
  kable_styling(font_size = 9)
```

Table 9: Average socio-economic characteristics of VV partisans

VV	Age	Female	Higher Education	Job
0	44.88	0.41	0.14	0.27
1	40.19	0.50	0.09	0.19
NA	42.75	0.57	0.13	0.28

Political Attitudes of VV Supporters in 2013

```
# find_var(ess_kosovo, pattern="trust")
# table(ess_kosovo$trstprl)
# table(ess_kosovo$trstprt)
kable(ess_kosovo  %>%
      group_by(vv) %>%
      select(trstprl, trstprt, trstep, vv) %>%
      summarise_all(funs(mean), na.rm = TRUE),
      caption = "Avgerage level of trust in political institutions (0 - 1
      digits=2, format = "latex", booktabs = T,
      col.names = c("VV", "Trust in Parliament", "Trust in Parties", "Tru
  kable_styling(font_size = 9)
```

Table 10: Avgerage level of trust in political institutions (0 - 10 scale)

VV	Trust in Parliament	Trust in Parties	Trust in EP
0	2.97	2.42	4.59
1	2.59	2.43	5.00
NA	2.50	1.82	4.12

References

Bowers, Jake, and Maarten Voors. 2016. "How to Improve Your Relationship with Your Future Self." *Revista de Ciencia Política* 36 (3). Pontificia Universidad Católica de Chile.