European Homocide Exercise

Rikke Uldbæk

07 September, 2022

Contents

The long-term trend in Homicides in Western Europe 1
Load the available data from ourworldindata.org 2
Inspect the data 2
Let's see what the long-term trend is in homicides 3
Uncouple the homicides of individual countries for easier view 3
Compare the trends in homicide with the pattern of reign duration among Danish rulers through time. 4
Final tasks 6

This exercise is based on the dataset provided by OurWorldInData project based at the Oxford University.

The long-term trend in Homicides in Western Europe

Understanding how homicide rates have changed prior to the modern era requires the help of historians and archivists. Manuel Eisner, a criminology professor at the University of Cambridge, and his colleagues published the Historical Violence Database : a compilation of data on long-term trends in homicide rates, in addition to qualitative information such as the cause of death, perpetrator and victim. This database is limited to countries with relatively complete historical records on violence and crime – mainly Western Europe and the US.

Starting in the second half of the nineteenth century, these European regions have consistent police records of those accused of murder or manslaughter and annual counts of homicide victims. To go back further in time, reaching as far back as the thirteenth century, Eisner collected estimates (from historical records of coroner reports, court trials, and the police) of homicide rates made in over ninety publications by scholars.

Homicide rates – *measured as the number of homicides per 100,000 individuals* – up to 1990 are sourced from Eisner's (2003) publication and the Historical Violence Database.

Are homicide rates in Europe today lower or higher than in the past? Using the provided dataset, display and describe the long-run homicide rates for the five European regions: Italy, England, Germany, Netherlands and Scandinavia.

```
#loading packages
pacman::p_load(tidyverse,dplyr, data.table, ggplot2, tibble, devtools, readr)

# setting working directory
setwd("~/Desktop/OneDrive - Aarhus Universitet/Cognitive Science/5th semester/Cultural Datascience/AU674
```

Load the available data from ourworldindata.org

You should always interrogate the source of your data. Who compiled it, from where, what is missing, how representative the data are? Check the data/Metadata.txt to learn about the data provenance.

```
Western_Europe <- read_csv("data/homicide-rates-across-western-europe.csv")
```

Inspect the data

How clean and analysis-ready is the dataset? Do you understand what the column names represent? What is the difference between rate and homicide number?

The column "Entity" represents the country or aggregated countries (e.g. Scandinavia) of measured homicide. The column "Code" represents a land code for simplicity of each of the elements in "Entity", however this one is mostly NA's. The column "Year" represents which year the homicide rate was measured. Lastly, the "Homicide Rate" represents the homicide rate in Europe over long-term per 100,000 individuals, i.e, number of homicides per 100,000 individuals. As an example, the first rows in the dataset that has a rate of 23, meaning, per 100,000 individuals 23 homicides were comitted.

```
head(Western_Europe)
```

```
## # A tibble: 6 x 4
##   Entity  Code   Year `Homicide rate in Europe over long-term (per 100,000) (h~`
##   <chr>   <chr> <dbl>                                                    <dbl>
## 1 England <NA>   1300                                                       23
## 2 England <NA>   1550                                                        7
## 3 England <NA>   1625                                                        6
## 4 England <NA>   1675                                                        4
## 5 England <NA>   1725                                                        2
## 6 England <NA>   1775                                                        1
```

Ok, the data look good except for the column `Homicide rate in Europe over long-term (per 100,000)` *which is not very easy to work with.*

- *Use the* `names()` *function and assignment key to relabel this column to* `homicides_per_100k`

```
# Checking the names
names(Western_Europe)
```

```
## [1] "Entity"
## [2] "Code"
## [3] "Year"
## [4] "Homicide rate in Europe over long-term (per 100,000) (homicides per 100,000 people)"
```

```
# Renaming the variable
Western_Europe<- rename(Western_Europe, "homicides_per_100k"= "Homicide rate in Europe over long-term (
```

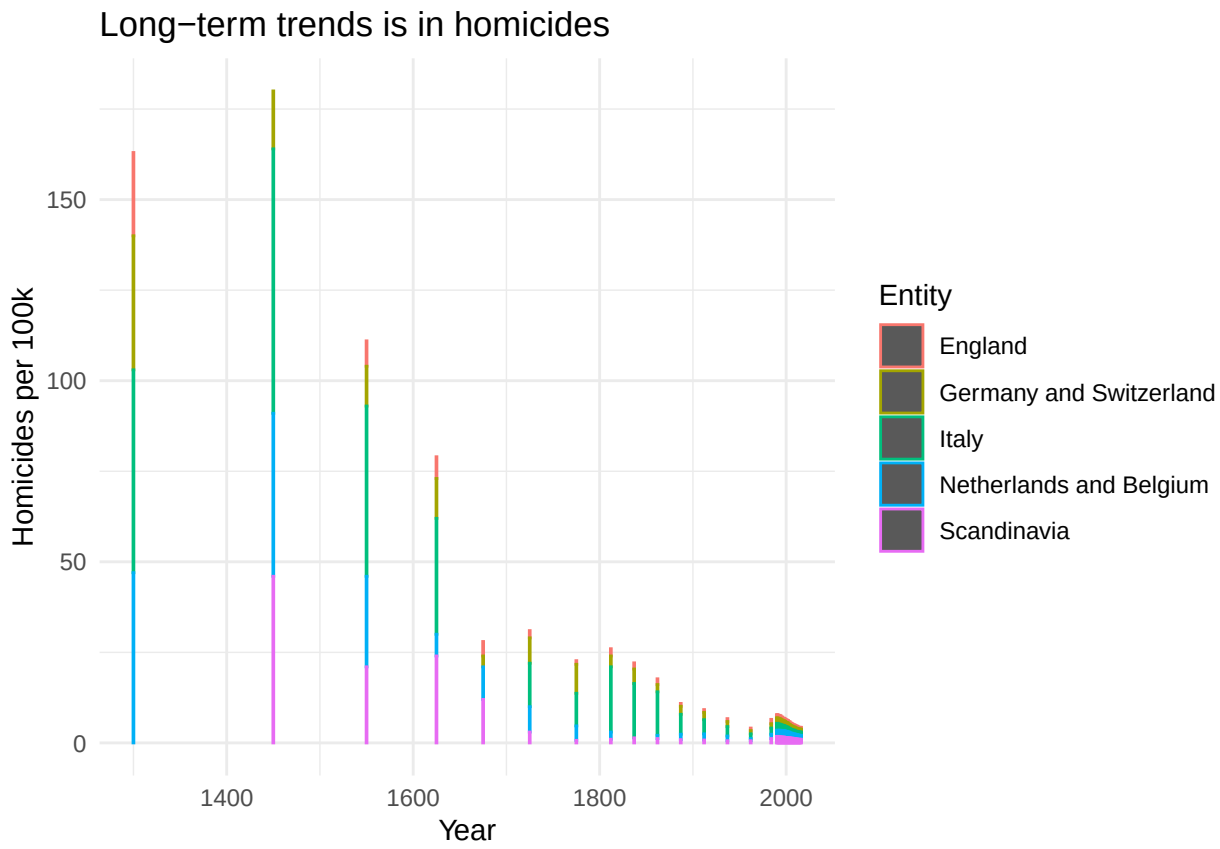
Now, that you have looked at what the data looks like and what it represents, and streamlined it, let's see what big picture it contains.

Let's see what the long-term trend is in homicides

- *use `ggplot()` function and remember the+ at the end of the line*
- *chose a meaningful `geom.....()` for geometry (hint: points are not great)*
- *load `Year` on the `x` axis and `homicides_per_100k` column in y axis*
- *to color individual country entries consistently, assign the country column to the argument `color`.*
- *provide meaningful title and axis labels*
- *remember to change the `eval` flag so that the code chunk renders when knitted*

```
#plotting the homicide trends
ggplot(data = Western_Europe, aes(x= Year, y= homicides_per_100k, color= Entity))+ geom_histogram(stat=
  labs(x = "Year", y = "Homicides per 100k")+
  ggtitle("Long-term trends is in homicides")+
  theme_minimal() #looks uninformative
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



Alright, the homicide rates should all be descending over time. What a comfort. But the viz is not super clear. Let's check the rates for individual countries.

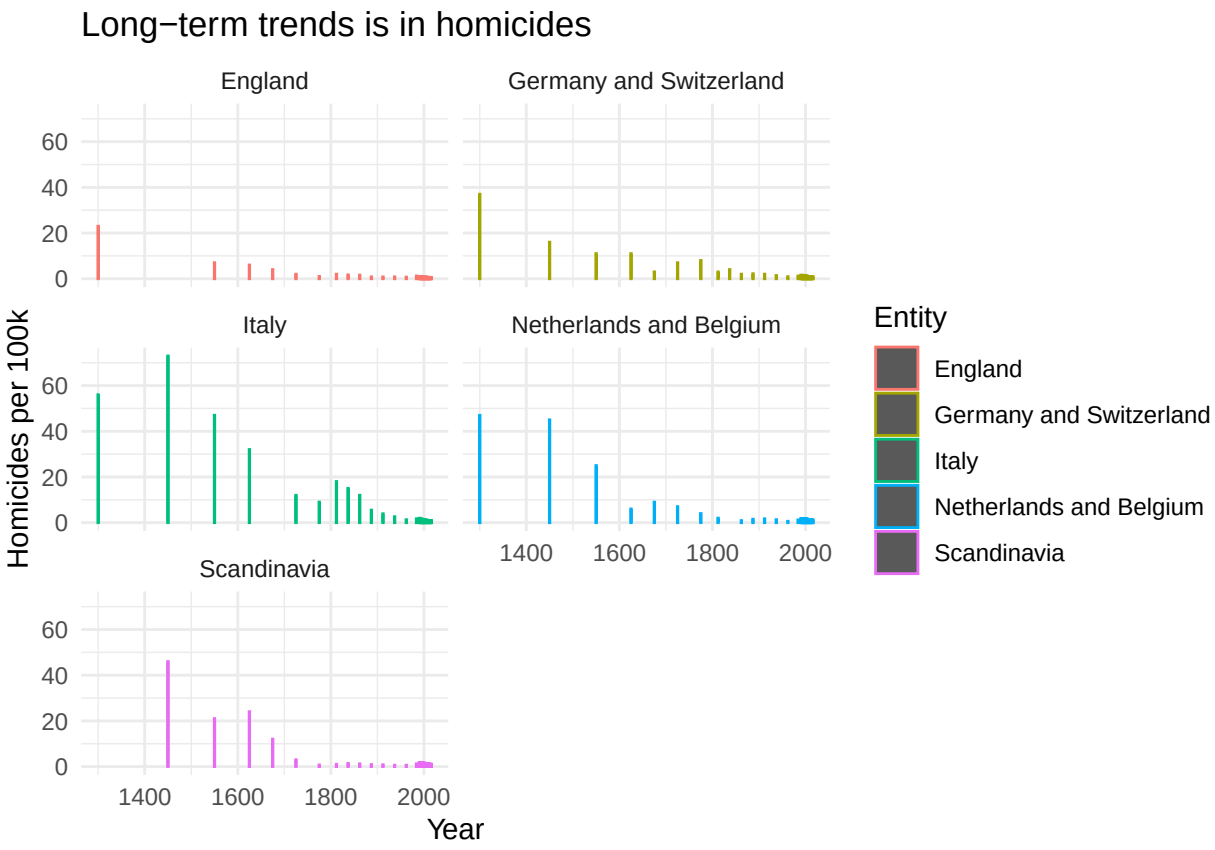
Uncouple the homicides of individual countries for easier view

You can visualize each country's trend separately by adding an extra argument to the ggplot, the `facet_wrap()` and feeding it the country column. If in doubt, check your ggplot tutorial and your country column name for exact usage.

- *reuse the ggplot from the chunk above*
- *insert `facet_wrap()` after the specification of geometry to split countries in separate charts*
- *change the facet "layout" to two columns and three rows so that the trends are easier to see in horizontal layout.*

```
#plotting the homicide trends
ggplot(data = Western_Europe, aes(x= Year, y= homicides_per_100k, color= Entity))+ geom_histogram(stat=
  facet_wrap(~Entity, nrow =3, ncol = 2)+
  labs(x = "Year", y = "Homicides per 100k")+
  ggtitle("Long-term trends is in homicides")+
  theme_minimal()
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



Compare the trends in homicide with the pattern of reign duration among Danish rulers through time.

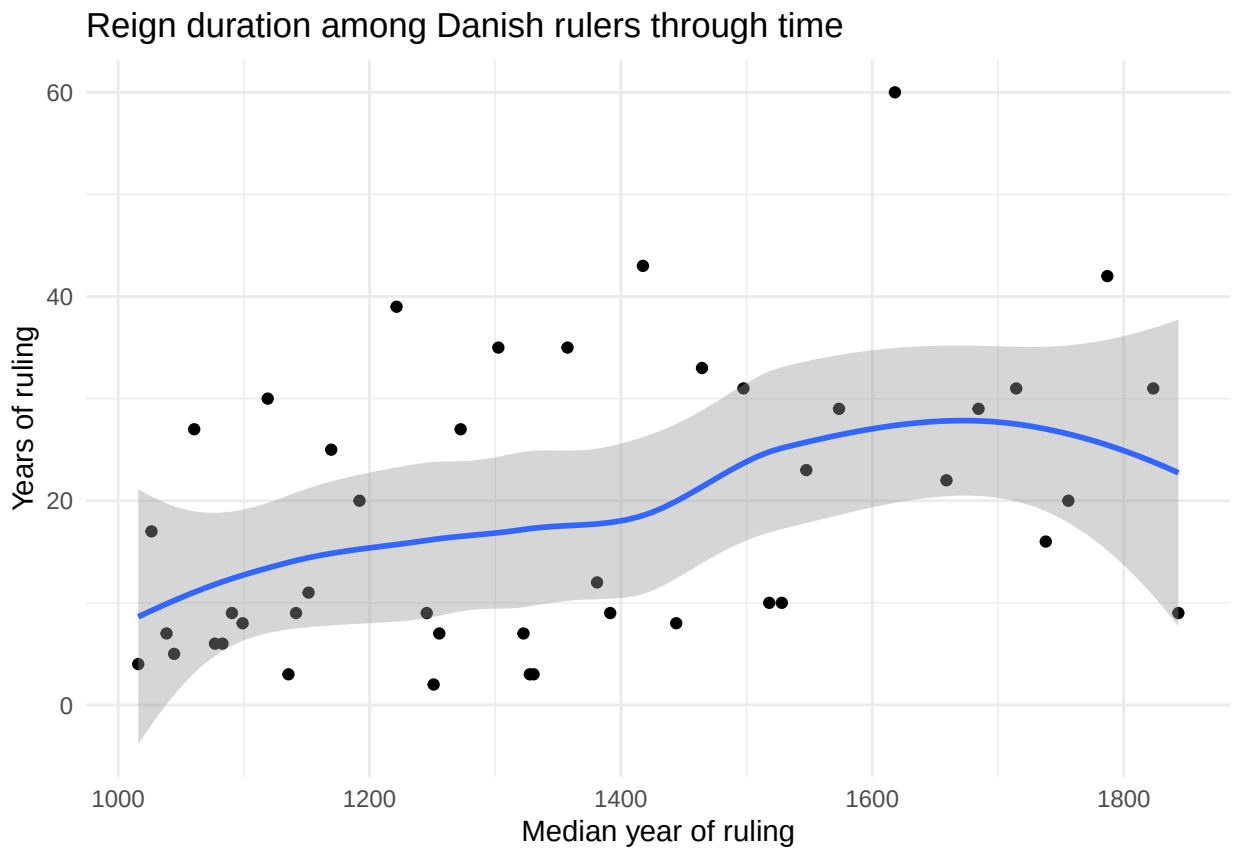
- *Load your Danish king dataset. Hopefully it is tidy and your years and duration of reign are all numeric.*
- *You need to have a consistent way of plotting the rulers' reign on the x axis, so I recommend you create a midyear column by calculating the middle of each monarch's rule (Hint: `midyear` = `endyear` - (`endyear-startyear`)/2)*
- *Start a ggplot plotting midyear on x axis and duration on y axis*
- *Try `geom_smooth()` for geometry*

- *Provide meaningful labels and a title*
- *How would you characterize the trend compared to the homicides above?* There seems to be an increase in "Years of ruling" through time, while the trend in the homicides above, yields a descending pattern in homicides through time.

```
# loading the data
kings <- read_delim("data/kings.csv", ";", escape_double = FALSE,
    trim_ws = TRUE)

#checking the data
ls.str(kings)
```

```
## End_date :  num [1:47] NA NA NA 1018 1035 ...
## Kings :  chr [1:47] "Gorm den Gamle" "Harald 1. Bl\xe5tand" "Svend 1. Tvesk\xe6g" ...
## Start_date :  num [1:47] NA NA NA 1014 1018 ...
## Yearasruler :  chr [1:47] "Unknown" "Unknown" "Unknown" "4" "17" "7" "5" "27" "6" "6" "9" ...
```

```
#Cleaning the data for good practice
kings <- na.omit(kings)
kings$Yearasruler <-  as.numeric(kings$Yearasruler)


#create a midyear column
kings <- kings %>%
  mutate(midyear= (End_date - (End_date - Start_date)/2))

#ggplot plotting midyear on x axis and duration on y axis
ggplot(kings,  aes(midyear, Yearasruler)) +
  geom_point()+
  geom_smooth()+
  labs(x = "Median year of ruling", y = "Years of ruling")+
  ggtitle("Reign duration among Danish rulers through time")+
  theme_minimal()
```




```
#checking variables for comparison
min(kings$midyear)
```

```
## [1] 1016
```

```
max(kings$midyear)
```

```
## [1] 1843.5
```

```
min(Western_Europe$Year)
```

```
## [1] 1300
```

```
max(Western_Europe$Year)
```

```
## [1] 2016
```

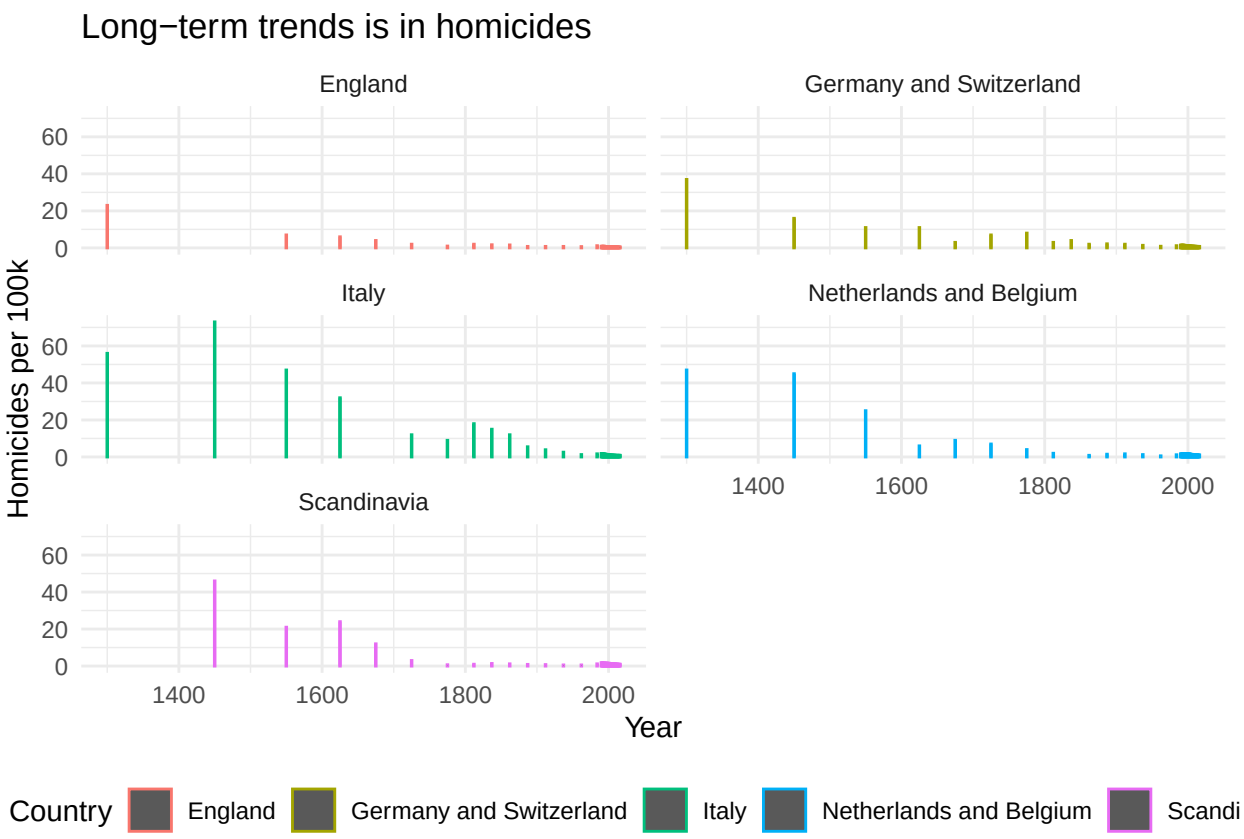
Final tasks

1) Plot: In the facetted plot above, move the legend from the current position on the side to below the facets, and label it "Country" instead of "Entity".

```
#plotting the homicide trends
p<- ggplot(data = Western_Europe, aes(x= Year, y= homicides_per_100k, color= Entity))+ geom_histogram(s
  facet_wrap(~Entity, nrow =3, ncol = 2)+
  labs(x = "Year", y = "Homicides per 100k")+
  ggtitle("Long-term trends is in homicides")+
  theme_minimal()
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

```
#Changing legend title and position
p + theme(legend.position="bottom") + guides(color=guide_legend(title="Country"))
```



2) Rmarkdown:

- *edit the author of the document, and convert 'Final Tasks' into heading #2 (like the other headings)*
- *add a* `floating table of contents` *to your Rmarkdown document*
- *provide informative* `chunk-names` *and edit flags in your R chunks, and automatically generate a* `timestamp` *to show when the document was last updated. (Hint: check the Rmarkdown episode in our Data Carpentry tutorial)*

3) Question: In <250 words articulate your answer on the basis of the data visualisations to the following question: are we more civilized today?

There seems to be an increase in "Years of ruling" through time, while the trend in the homicides above, yields a descending pattern in homicides through time. However, before starting to compare the trends to

much, we need to consider that the time lines does not overlap completely. The earliest year a ruler started ruling was in 1016 (considering the midyear variable), and the earliest year we have data about homicides are from 1300, meaning we have a gap of 300 years, where we lack homicide data, thus making parts of this exercise's graph not comparable. Same goes for the later years, where the latest year reported of a ruler was 1843.5 (considering the midyear variable), while the latest year we have data about homicides are from 2016, meaning we have a gap of 172.5 years lacking ruler-data. Lastly, in order to compare such trends, simply just eyeballing the data is insufficient. A statistical test would be needed to make a proper comparison. Thus, from these data and plots, I will not imply any correlation before a statistical test is conducted. This stated, we should hopefully be more civilized today.