194.044 Data Stewardship

Assignment 1

Aleksandar Pavlovic

16.04.2019

Question

This document deals with answering the pseudo question, if the global temperature trend influences the number of children born alive in Vienna.

Datasets

To answer the proposed question above, the following two datasets are used:

- Births in Vienna - identifier: f54e6828-3d75-4a82-89cb-23c58057bad4, url: [birth data repo](https://www.data.gv.at/katalo 3d75-4a82-89cb-23c58057bad4)
- Mean near surface temperature deviation (the first zip file was downloaded) - identifier: *cli_iad_td*, url: [temperature repo](https://data.europa.eu/euodp/en/data/dataset/zQAEvhkR7H0NQYU1HP5fA)

```
# Load libraries
library(ggplot2)
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.5.3
```

```
## -- Attaching packages ----------------------------------------------------------------------
```

```
## v tibble  1.4.2     v purrr   0.2.5
## v tidyr   0.8.3     v dplyr   0.7.7
## v readr   1.1.1     v stringr 1.3.1
## v tibble  1.4.2     v forcats 0.3.0
```

```
## Warning: package 'tidyr' was built under R version 3.5.3
```

```
## -- Conflicts -------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

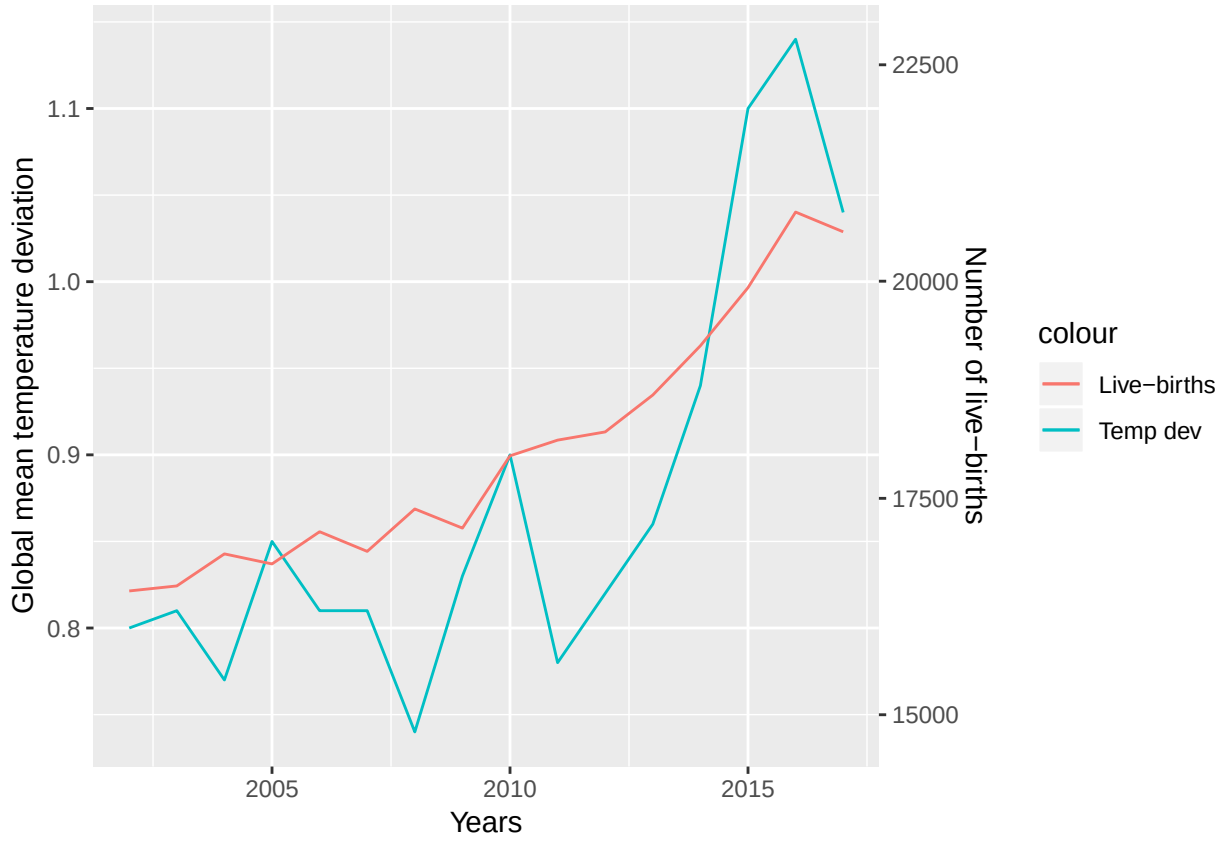
```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 3.5.3
```

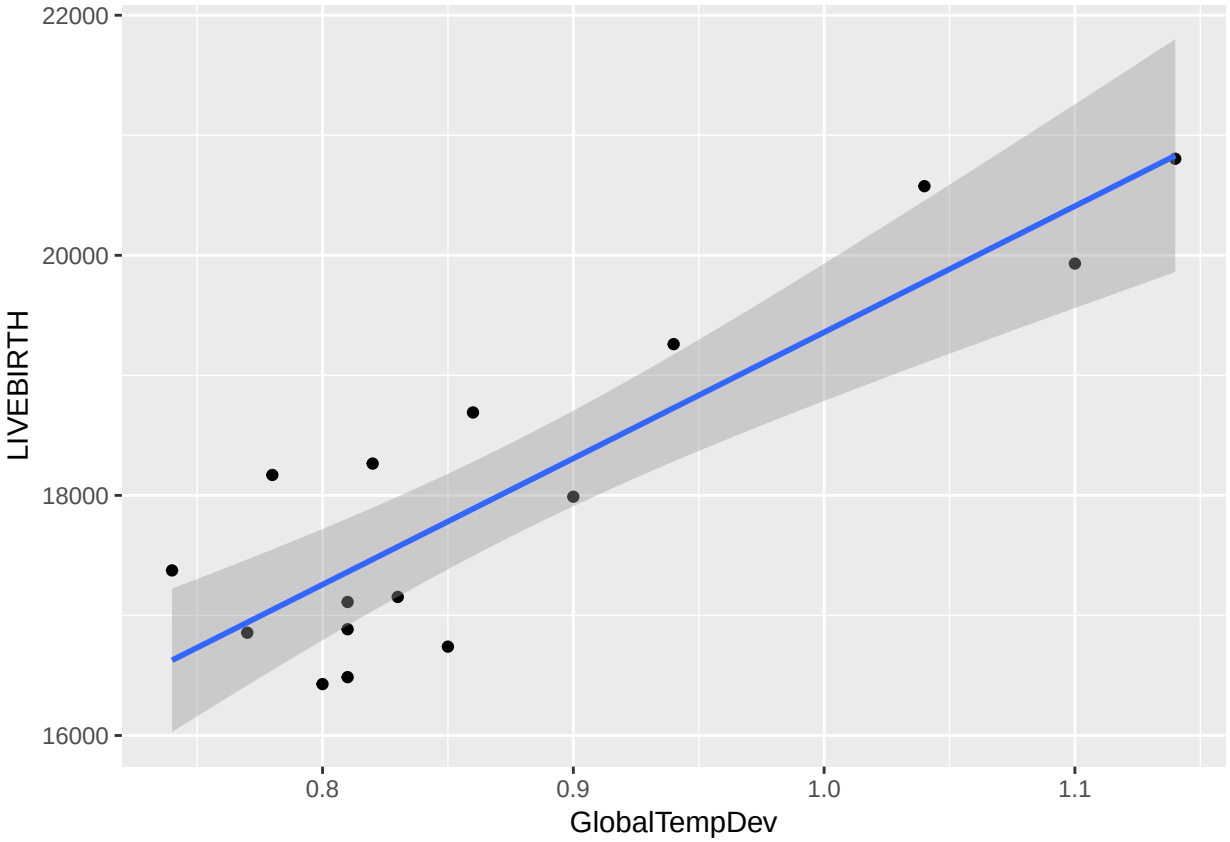
```
## corrplot 0.84 loaded
```

```
# Load preprocessed and merged data
data <- read.csv2("./data/processed/mergedData.csv")
```

```
# First visualisation:
ggplot(data=data, aes(x=year, group=1)) +
  labs(y = "Global mean temperature deviation",
                x = "Years") +
  geom_line(aes(y = GlobalTempDev, colour = "Temp dev")) +
  geom_line(aes(y = LIVEBIRTH/20000, colour = "Live-births")) +
  scale_y_continuous(sec.axis = sec_axis(~.*20000, name = "Number of live-births"))
```




```
# Second visualisation:
ggplot(data=data, aes(x=GlobalTempDev, group=1)) +
  geom_point(aes(y = LIVEBIRTH)) +
  geom_smooth(mapping = aes(x = GlobalTempDev, y = LIVEBIRTH), method=lm)
```


22000
20000
18000
16000
LIVEBIRTH
0.8
0.9
1.0
1.1
GlobalTempDev