Modeling global warming

Babich Nikita

2023-06-26

1 Modeling global warming

1.1 Introduction

Global warming is the biggest phenomenon which has yet unknown consequences on humanity. That is why I am interested whether it is possible to observe the evidence of the global warming with just temperature analysis (without research on humidity/atmospheric pressure etc.).

1.2 Data

From the website Kaggle I've downloaded a sheet with country-date-temperature relations. In this dataset we have a lot of months for which AverageTemperature is not available, so I've decided to delete these rows to speed up any future data manipulations. Also the first column has to be converted to time format. From the summary we can see that dataset contains around half a million observations. The time period of 1743-2013 overlaps with industrial and technical revolutions (although in this work I will not look into whether they are related to the data).

```
data <- read.csv("GlobalLandTemperaturesByCountry.csv")
dataNoNA <- data[!is.na(data$AverageTemperature), ]
dataNoNA$dt <- as.Date(dataNoNA$dt)
dataset_summary <- summary(dataNoNA)
dataset_summary
```

```
##        dt             AverageTemperature AverageTemperatureUncertainty
##  Min.   :1743-11-01   Min.   :-37.66     Min.   : 0.052
##  1st Qu.:1869-11-01   1st Qu.: 10.03     1st Qu.: 0.323
##  Median :1919-08-01   Median : 20.90     Median : 0.571
##  Mean   :1913-08-08   Mean   : 17.19     Mean   : 1.019
##  3rd Qu.:1966-10-01   3rd Qu.: 25.81     3rd Qu.: 1.207
##  Max.   :2013-09-01   Max.   : 38.84     Max.   :15.003
##    Country
##  Length:544811
##  Class :character
##  Mode  :character
##
##
##
```

1.3 Methodology

If we try to immediately plot temperature~time, then we get the following picture:

```
library(modelr)
```

```
## Warning: пакет 'modelr' был собран под R версии 4.3.1
```

```
mod1 <- lm(formula = AverageTemperature ~ dt, data = dataNoNA)
grid <- dataNoNA |>
  modelr::data_grid(dt) |>
  modelr::gather_predictions(mod1)

library(ggplot2)
```

```
## Warning: пакет 'ggplot2' был собран под R версии 4.3.1
```

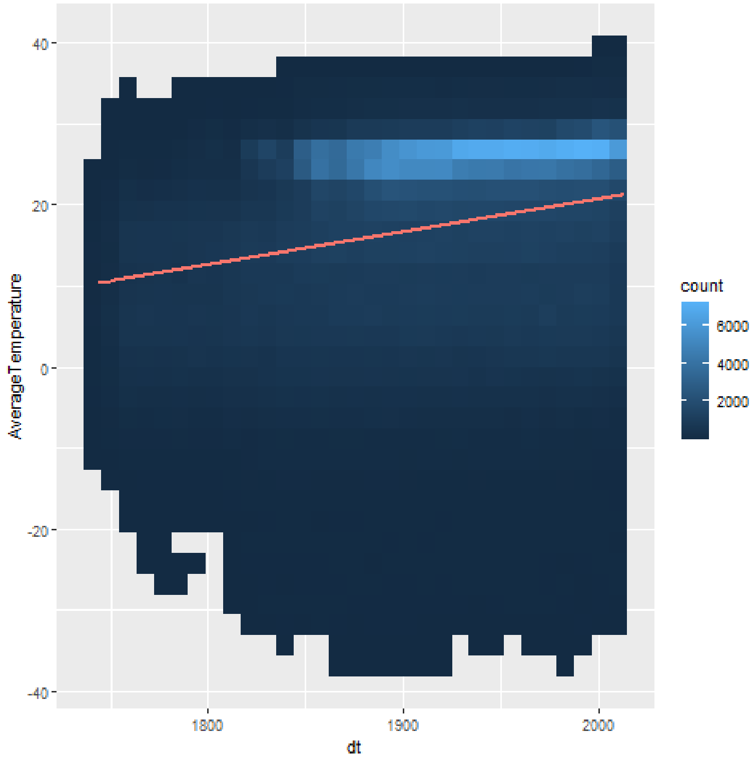
```
g <- ggplot(data = dataNoNA, mapping = aes(x = dt, y = AverageTemperature))
g +
  geom_bin2d() +
  #stat_summary(geom = "line", fun = median, aes(group = as.Date(dt, "%Y"))) +
  geom_line(data = grid, aes(y = pred, color = "red"), linewidth = 1, show.legend = FALSE)
```

```
dev.copy(png, "my_plot.png")
```

```
## png
##   3
```

```
dev.off()
```

```
## png
##   2
```



Relation between year and average temperature

As shown in Figure linear model has a positive slope, but we cannot make conclusions, because that shift could be caused by some countries with hot climate starting gathering data later than others. But if we plot data from within one country we will get the needed consistency. The historical changes in countries' area are usually negligible compared to their total area. Plotting for each of 242 countries would be problematic, so I will concentrate attention on countries with the biggest area. Just to cover all parts of the world I will add to the list couple countries: France and South Africa.

- Russia - 16,378,410 square kilometers
- Canada - 9,093,507 square kilometers
- China - 9,326,410 square kilometers
- United States - 9,147,593 square kilometers
- Brazil - 8,460,415 square kilometers
- Australia - 7,633,565 square kilometers
- India - 2,973,190 square kilometers
- Argentina - 2,736,690 square kilometers
- Kazakhstan - 2,699,700 square kilometers
- Algeria - 2,381,741 square kilometers
- France - 640,427 square kilometers
- South Africa - 1,214,470 square kilometers

```r
names <- c("Russia",
         "Canada",
         "China",
         "United States",
         "Brazil",
         "Australia",
         "India",
         "Argentina",
         "Kazakhstan",
         "Algeria",
         "France",
         "South Africa")
areas <- as.numeric(gsub(",", "", c("16,378,410",
                                    "9,093,507",
                                    "9,326,410",
                                    "9,147,593",
                                    "8,460,415",
                                    "7,633,565",
                                    "2,973,190",
                                    "2,736,690",
                                    "2,699,700",
                                    "2,381,741",
                                    "640,427",
                                    "1,214,470")))
info <- data.frame(names, areas) #, coef1 = double(length(names)), coef2 = double(length(names)))
info
```

```
##           names    areas
## 1        Russia 16378410
## 2        Canada  9093507
## 3         China  9326410
## 4  United States  9147593
## 5        Brazil  8460415
## 6     Australia  7633565
## 7         India  2973190
## 8     Argentina  2736690
## 9    Kazakhstan  2699700
## 10      Algeria  2381741
## 11       France   640427
## 12  South Africa  1214470
```

Areas of countries can be used to calculate the weighted average. Second coef allows to get information not only about temperature growth, but also about acceleration/deceleration, so I will use parabolic relation for the model. Acceleration would be presented by convex lines, deceleration - by concave. The coefficient from the model will be added to coef table.

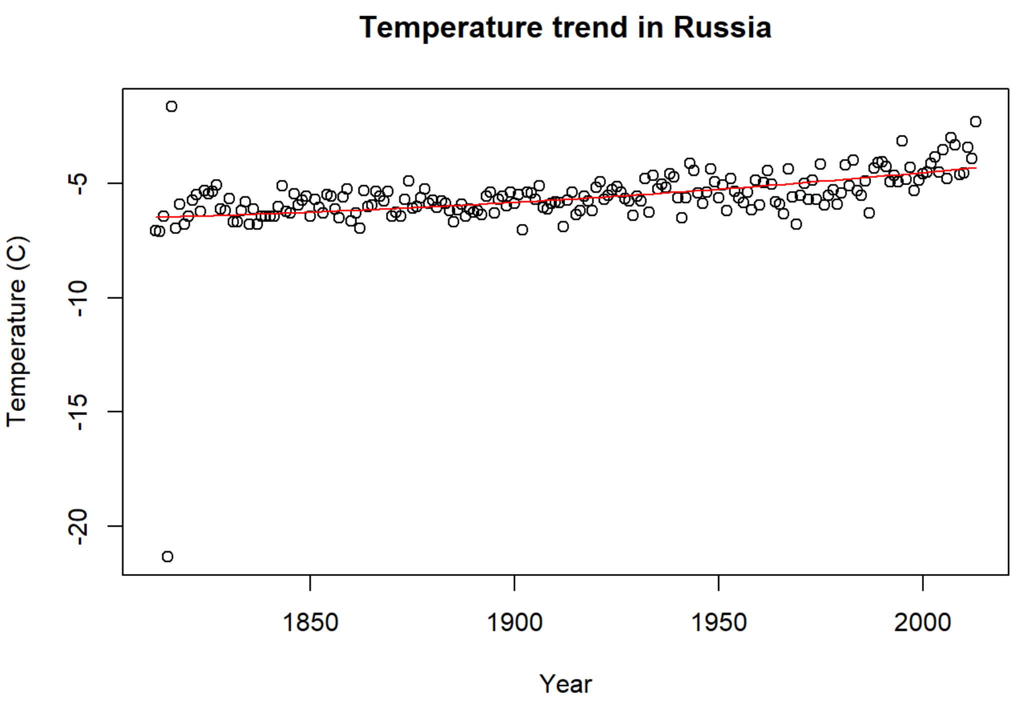
```r
library(dplyr)
```

```
## 
## Присоединяю пакет: 'dplyr'
```

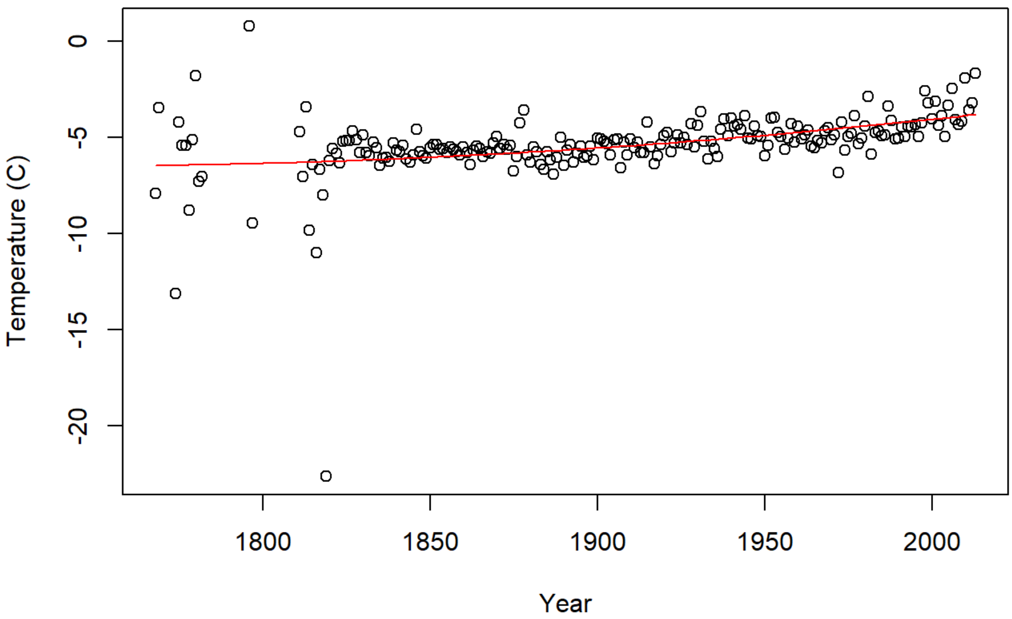
```
## Следующие объекты скрыты от 'package:stats':
## 
##     filter, lag
```

```
## Следующие объекты скрыты от 'package:base':
## 
##     intersect, setdiff, setequal, union
```

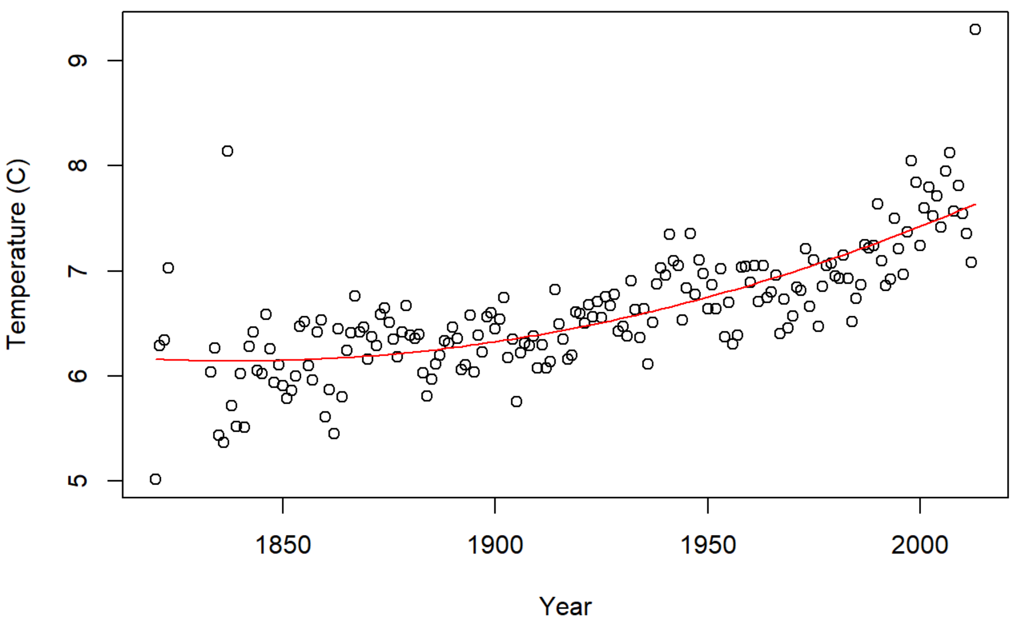
```
dfc <- data.frame(matrix(ncol = 3, nrow = 0)) #new dataframe for coefficients


dataNoNA <- dataNoNA %>%
    mutate(dt = as.integer(substring(dt, 1, 4)))

plotcountry <- function (countryname){
  chosen_data <- dataNoNA %>%
    filter(dataNoNA$Country == countryname)
  chosen_data <- chosen_data  %>%
    group_by(dt) %>%
    summarise(AverageTemperature = mean(AverageTemperature))

  #print(chosen_data)

  model <- lm(formula = AverageTemperature ~ poly(dt, 2), data = chosen_data)
  #print(model)
  dfc <<- rbind(dfc, c(countryname, as.numeric(coef(model)[2]) , as.numeric(coef(model)[3]) ))

  data_to_plot <- data.frame(dt = seq(min(chosen_data$dt),
                                      max(chosen_data$dt),
                                      length.out = 100))
  data_to_plot <- add_predictions(data_to_plot, model)


  plot(x = chosen_data$dt,
       y = chosen_data$AverageTemperature,
       ylab = "Temperature (C)",
       xlab = "Year",
       main = paste("Temperature trend in", countryname))
  lines(x = data_to_plot$dt,
        y = data_to_plot$pred,
        col = "red")

}

for(member in info$names){plotcountry(member)}
```

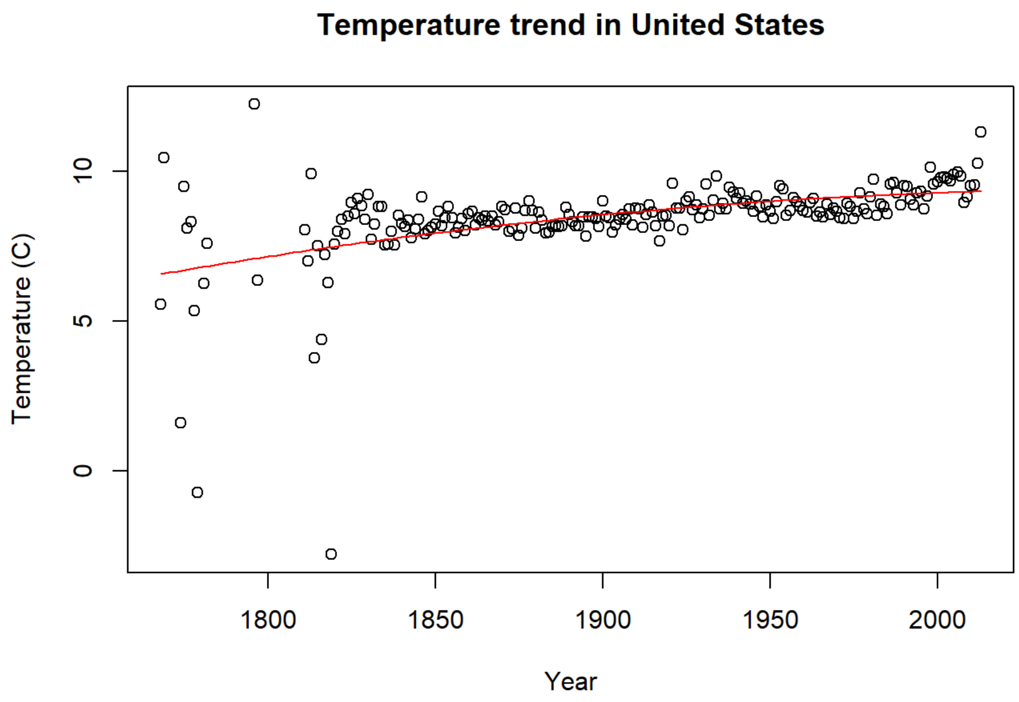


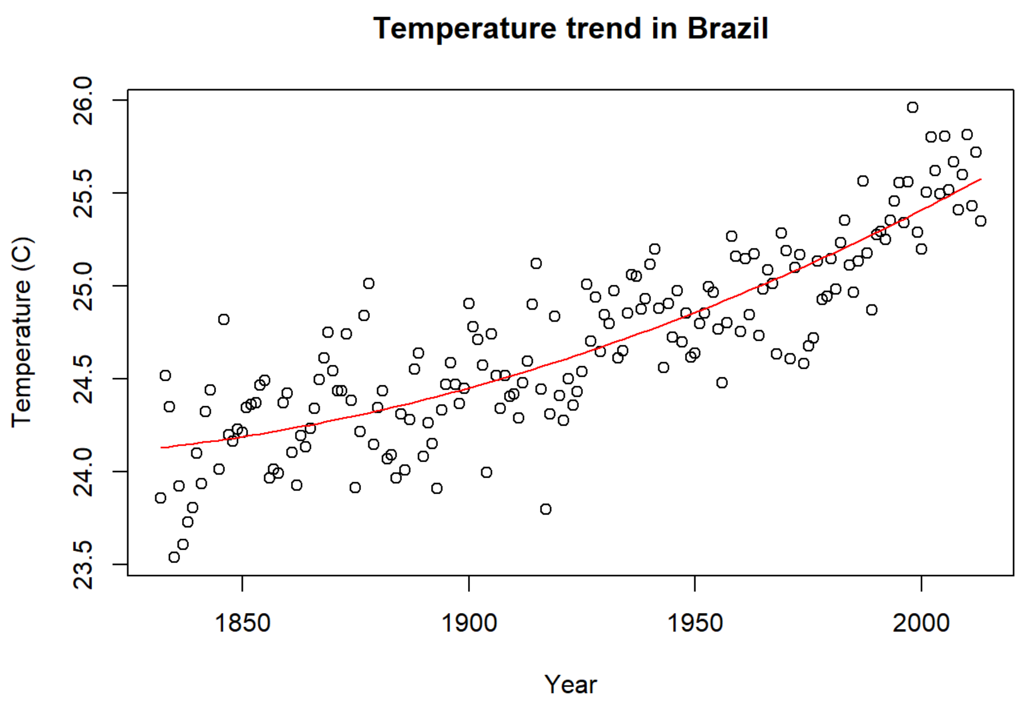
Temperature trend in Canada



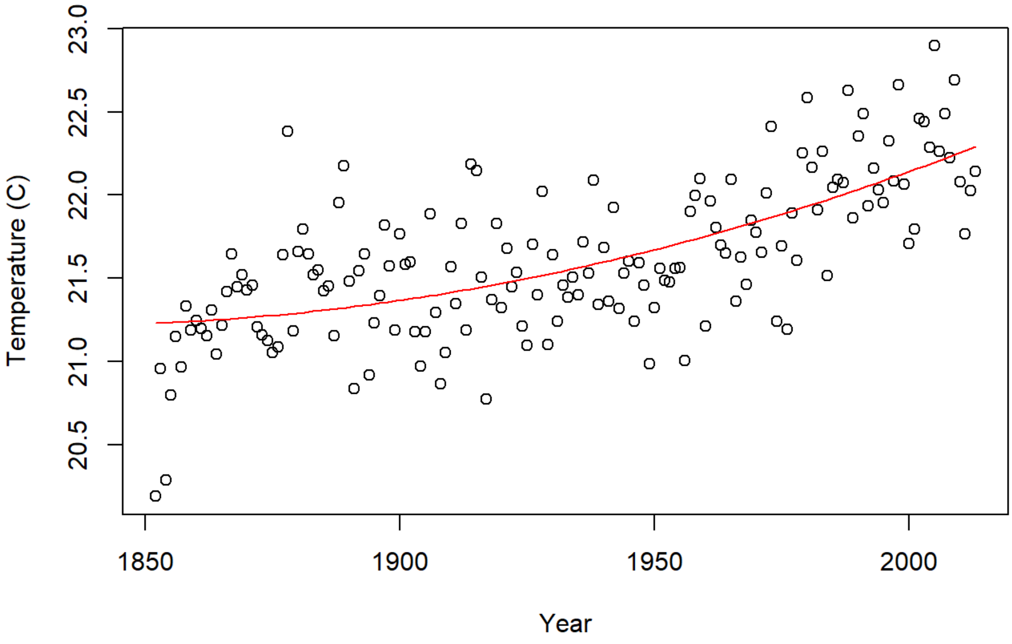
Temperature trend in China



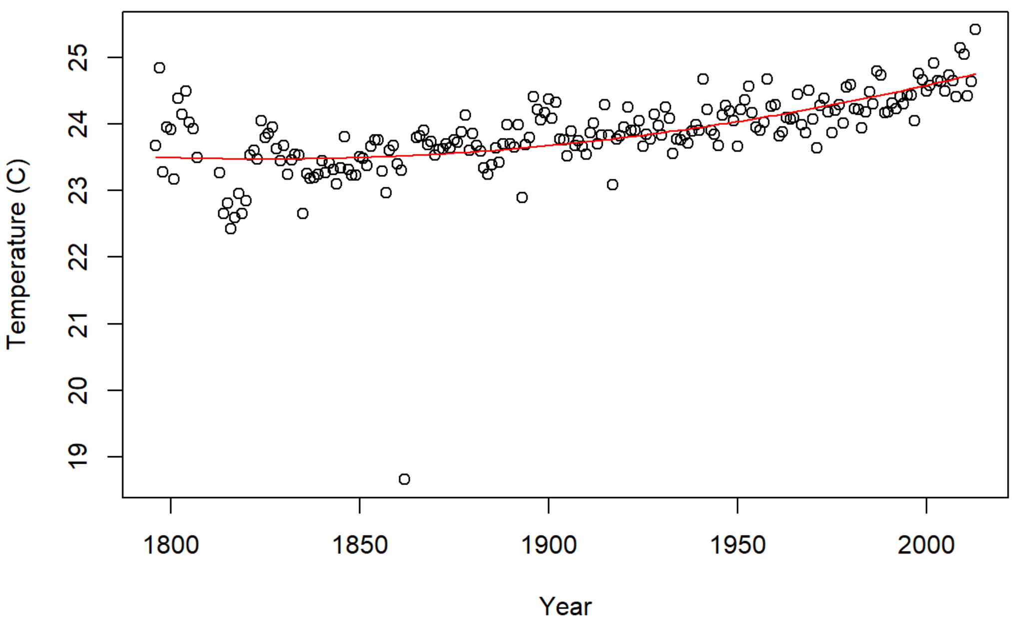

Temperature trend in United States
Temperature (C)
0
5
10
1800
1850
1900
1950
2000
Year


Temperature trend in Brazil
Temperature (C)
23.5
24.0
24.5
25.0
25.5
26.0
1850
1900
1950
2000
Year

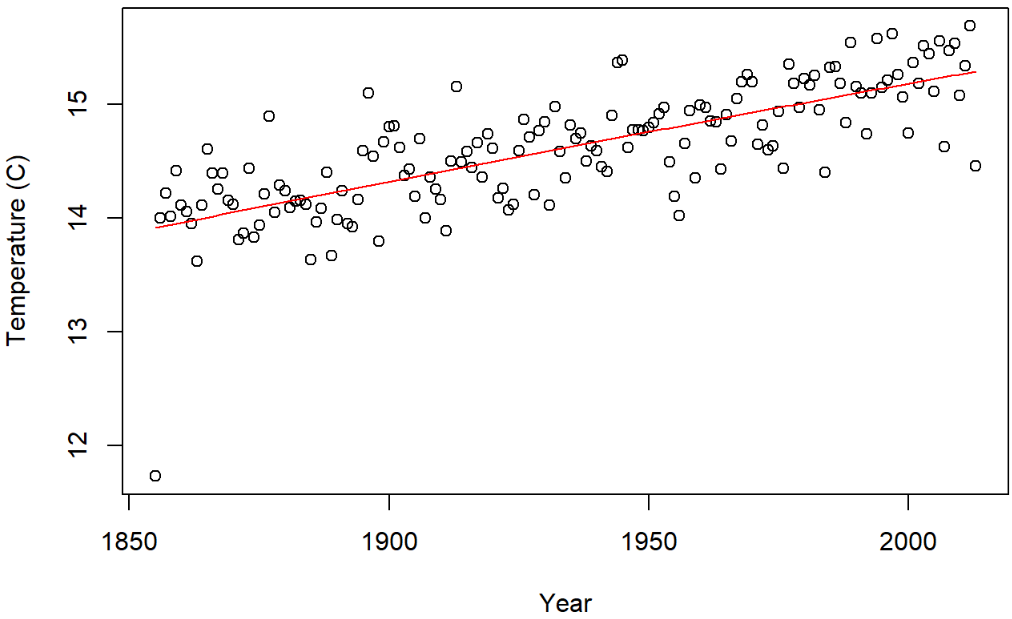
Temperature trend in Australia



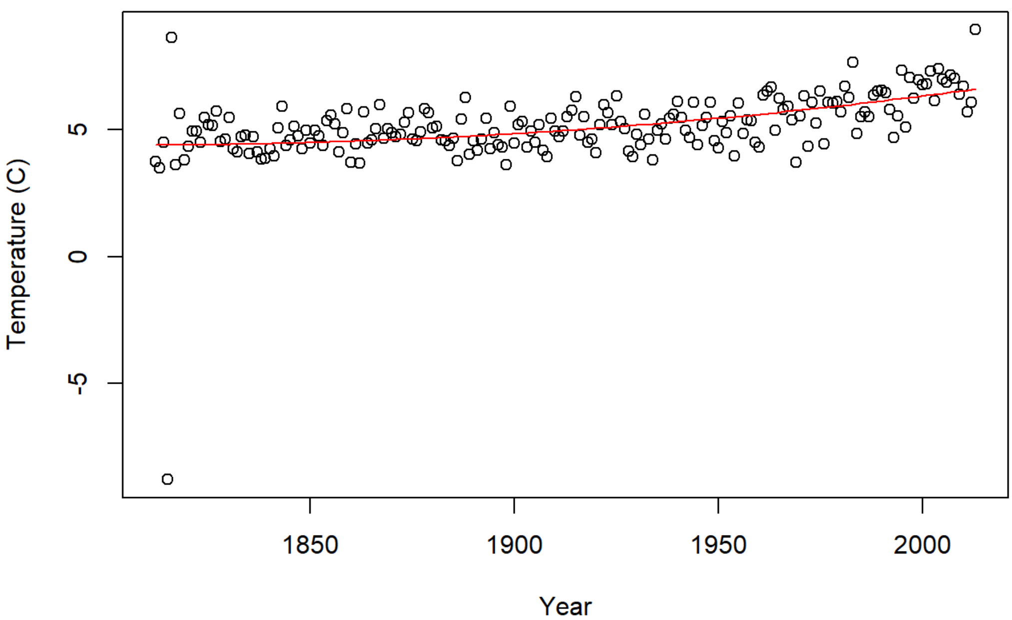
Temperature trend in India



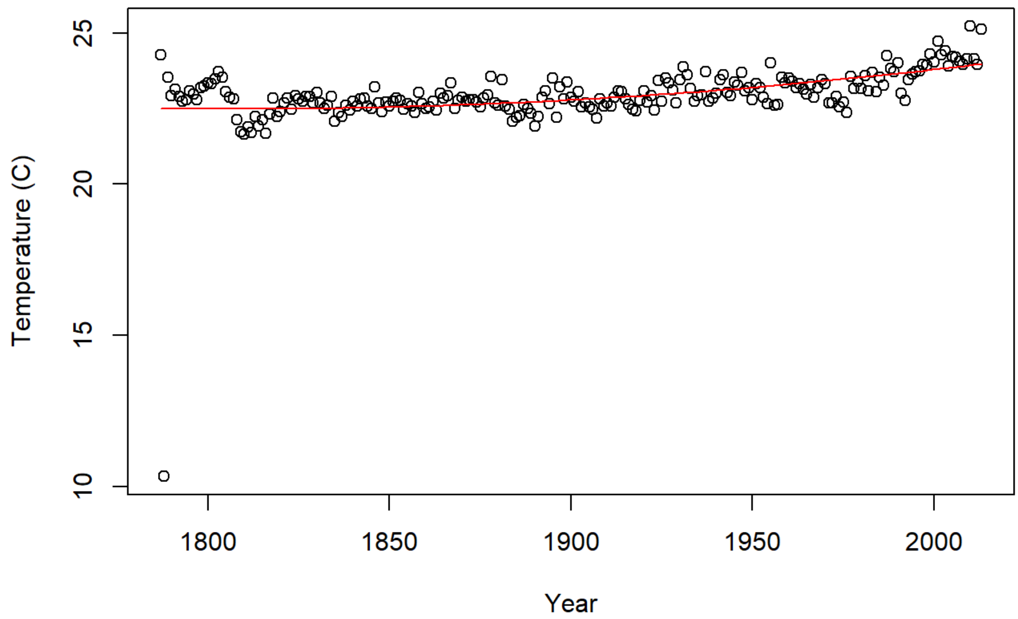
Temperature trend in Argentina



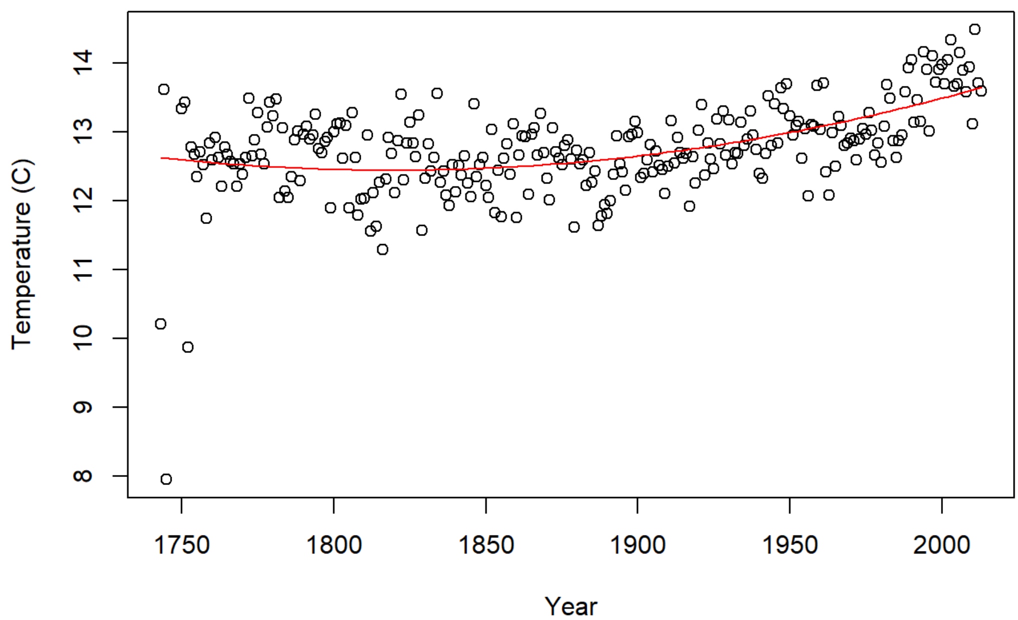
Temperature trend in Kazakhstan

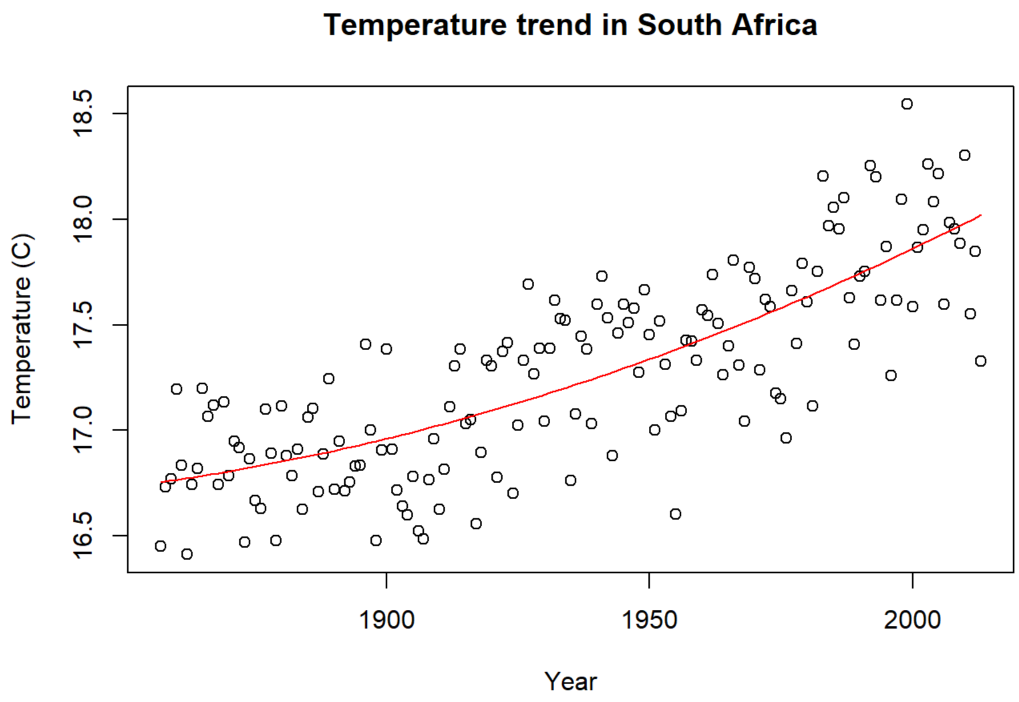


Temperature trend in Algeria



Temperature trend in France






```
print(info)
```

```
##          names    areas
## 1       Russia 16378410
## 2       Canada  9093507
## 3        China  9326410
## 4  United States  9147593
## 5       Brazil  8460415
## 6    Australia  7633565
## 7        India  2973190
## 8    Argentina  2736690
## 9   Kazakhstan  2699700
## 10     Algeria  2381741
## 11      France   640427
## 12 South Africa  1214470
```

```
colnames(dfc) <- c("country","coef1","coef2")
dfc <- dfc %>% mutate(coef1 = as.numeric(coef1),
       coef2 = as.numeric(coef2))
print(dfc)
```

```
##        country     coef1       coef2
## 1       Russia  8.992206  1.28642109
## 2       Canada 10.949737  2.14564092
## 3        China  5.867912  1.77262275
## 4  United States 10.049066 -2.04931096
## 5       Brazil  5.648817  0.95895348
## 6    Australia  3.920602  0.82090363
## 7        India  5.278786  1.84981238
## 8    Argentina  5.038068 -0.06416186
## 9   Kazakhstan  9.083308  2.23232781
## 10     Algeria  6.443524  2.06214978
## 11      France  4.900797  2.79908378
## 12 South Africa  4.610826  0.67631766
```

From the data we can see that only USA and Argentina have negative second coefficient, which means slowing down of the warming tempo, but all observed countries have a positive first coefficient.

```
total_land_area <- 148326000
total_area_countries <- sum(info$areas)
weighted_coef1 <- weighted.mean(dfc$coef1, w = info$areas)
weighted_coef2 <- weighted.mean(dfc$coef2, w = info$areas)
weighted_coef1
```

```
## [1] 7.557291
```

```
weighted_coef2
```

```
## [1] 0.985371
```

```
total_area_countries/total_land_area
```

```
## [1] 0.490043
```

1.4 Conclusion

In this project it is shown that it is possible to observe global warming phenomena from just temperature data. This work proves that for almost 50% of Earth land surface the temperatures are growing, and that for most of that territory even acceleration takes place.

2 Potential continuation

General sigmoid function

$$\sigma(x) = \frac{k4}{1 + e^{-(x * k3 + k1)}} + k2$$

could be used as a non-linear model with k1, k2, k3, k4 unknowns, and it could predict the level at which the temperature stabilizes after the global warming. But it has couple limitations: - 1) the slow-down has to have begun to happen, or else it would be an attempt to approximate exponential/polynomial/linear function with a sigmoid, which has an upper limit. That will send that limit arbitrarily high up to allow the best fit with the left part of the sigmoid. But if it coef2 becomes negative, meaning the new observations form a concave function, then sigmoid will be applicable. - 2) it would be easier to find all 4 parameters with gradient descend rather than with projection method of finding approximations.

3 References

1. Bacigál, T. (2022). Úvod do analýzy údajov pomocou R. Spektrum STU. Link

2. Sissener, K. (2017). Climate Change: Earth Surface Temperature Data. Link

3. List of countries and dependencies by area. Link