#Лабораторный практикум c R

**Преподаватель**

Дорошко Ольга Валерьевна

**Студент**

Клюев Владислав Дмитриевич, 1 курс, Аналитическая логистика

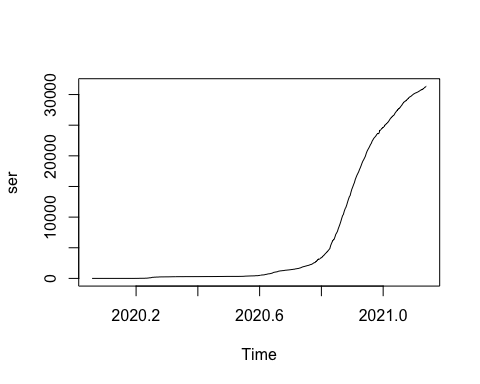
# Задание 2. Прогнозирование эпидемиологической ситуации

Это задача прогнозирования временных рядов. Сначала считаем данные, выберем количество случаев в провинции Манитоба в Канаде, выберем даты выборки данных и переведем в формат time series (ts), проиллюстрируем наглядно на графике.

library(forecast)

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

library(tseries)  
options(warn=-1)  
  
data <- read.csv2('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse\_covid\_19\_data/csse\_covid\_19\_time\_series/time\_series\_covid19\_confirmed\_global.csv',sep=',',stringsAsFactors = FALSE,header = TRUE)  
data = data[data$Province.State == "Manitoba",]  
data = unlist(data, use.names = FALSE)  
data = as.numeric(data)  
data = data[5:length(data)]  
idx <- seq(as.Date("2020-01-22"), Sys.Date()-1, by = "day")  
ser <- ts(data, start = c(2020, as.numeric(format(idx[1], "%j"))), frequency = 365)  
plot.ts(ser)



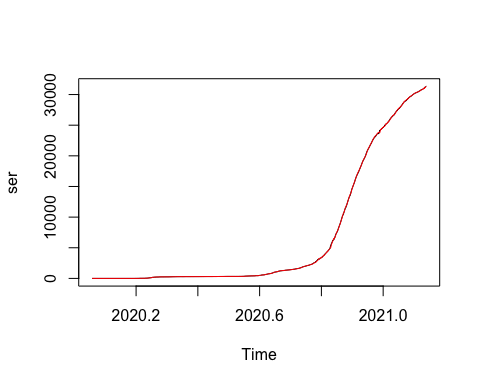
Для построения прогноза будем использовать модель типа ARIMA. Конечно, можно было это сделать вручную, сначала с помощью дифференцирования приведя ряд к стационарному (функция diff(data, differences=x)) и проверив его на стационарность с помощью теста Дики-Фулера (adf.test(data)), подобрать p и q для модели, с помощью acf() и pcf(), но если можно это сделать автоматически, то так и поступим.

auto.arima(ser)

## Series: ser   
## ARIMA(3,2,2)   
##   
## Coefficients:  
## ar1 ar2 ar3 ma1 ma2  
## 0.9048 -0.1847 0.1483 -1.7645 0.8240  
## s.e. 0.0646 0.0673 0.0555 0.0451 0.0391  
##   
## sigma^2 estimated as 2282: log likelihood=-2080.79  
## AIC=4173.58 AICc=4173.8 BIC=4197.44

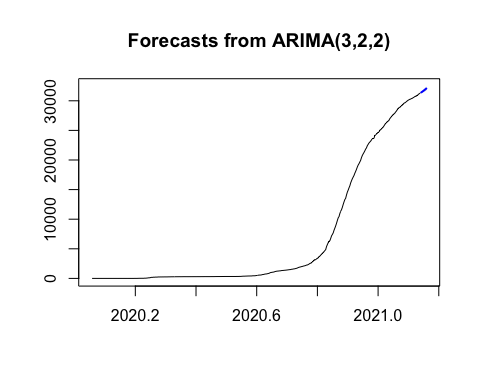
Чудесно, теперь построим модель ARIMA, используя автоматически найденные коэффициенты. Построим график исходных данных и полученных в результате предсказания.

serarima <- arima(ser, order=c(3,2,2))  
plot(ser)  
lines(fitted(serarima), col='red')



Построим предсказание числа заболевших на ближайшие 7 дней с уровнем уверенности 95% для доверительного интервала (я не совсем знаю, как это правильно сказать на русском: confidence level of 95% for prediction intervals). Он также даст точечное предсказание. Изобразим на графике.

serarimaforecast <- forecast(serarima, h=7, level=c(95))  
plot(serarimaforecast)



serarimaforecast

## Point Forecast Lo 95 Hi 95  
## 2021.1425 31421.96 31328.94 31514.99  
## 2021.1452 31519.64 31378.56 31660.73  
## 2021.1479 31622.38 31445.78 31798.97  
## 2021.1507 31728.66 31513.30 31944.02  
## 2021.1534 31837.91 31577.65 32098.17  
## 2021.1562 31949.95 31639.52 32260.38  
## 2021.1589 32064.49 31698.46 32430.52