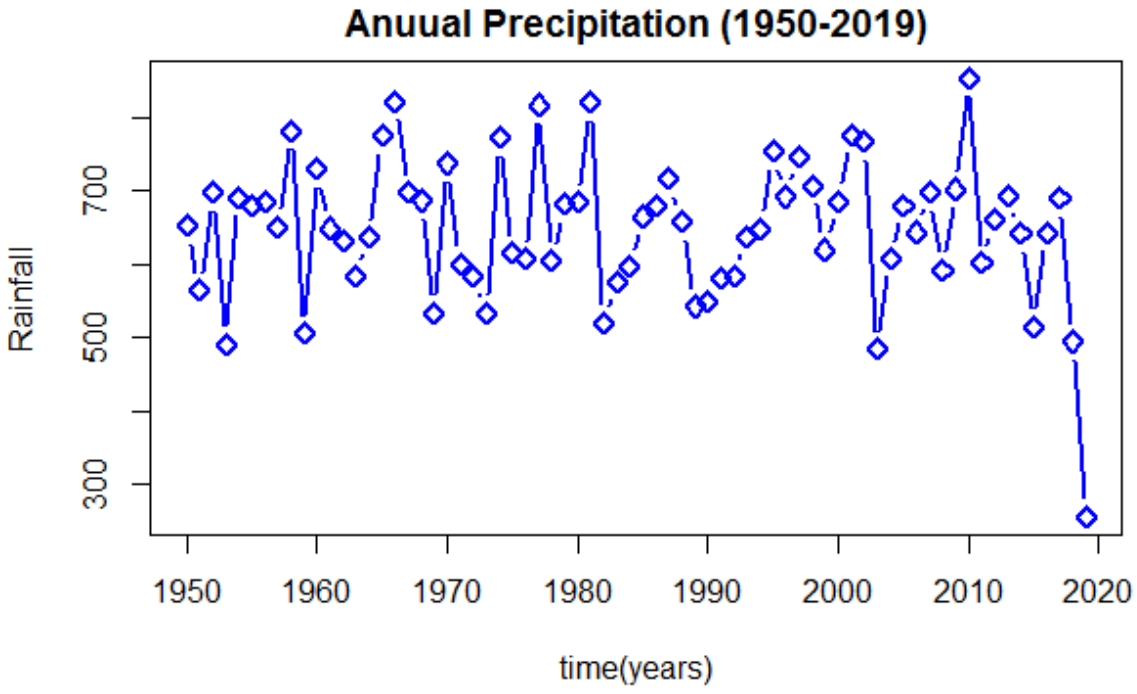
DATA ANALYSIS OF CZECH REPUBLIC RAINFALL 1950-2019

Sadaf Nasreen (22/08/2019)

This script solves the following contents,

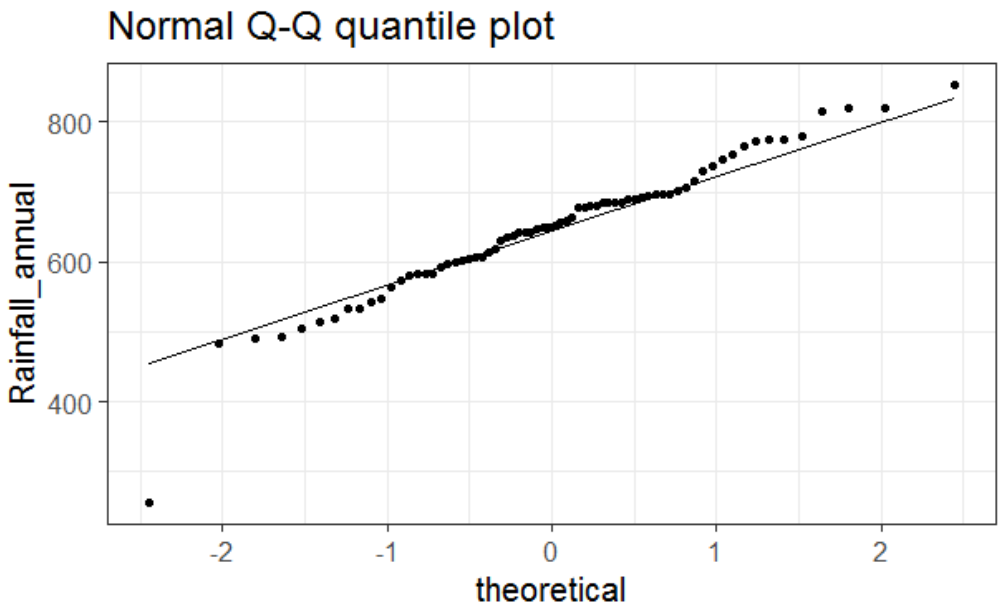
- First prepare the data set to convert daily dataset into months and annual using sum function
- Draw line and quantile plot
- Draw trend series of annual data set
- Draw mean monthly precipitation using bar plot
- Draw sum of monthly precipitation using box plot
- Trend and significance value of monthly rainfall

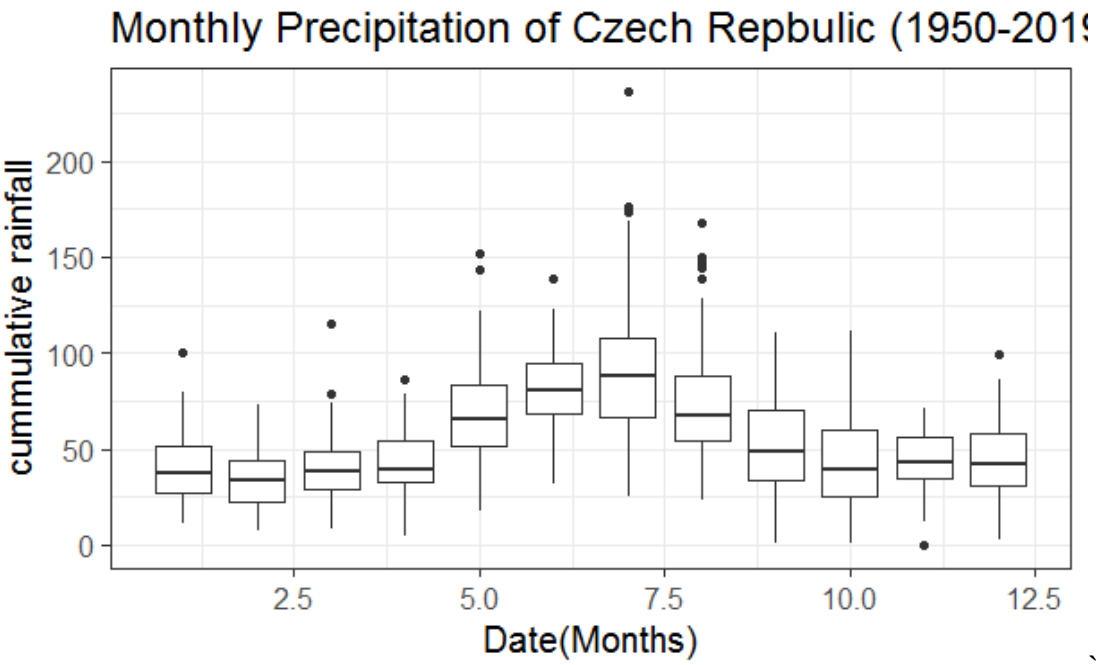
The Rainfall_script.R is an explanatory code that calculates annual precipitation amount using line and quantile plots. Blue color shows the annual variation of rainfall at Czech republic starts from January 1950 to May 2019, goes to sudden decrease in rainfall at year 2019.



The Q-Q plot, or quantile to quantile plot, is a graph that tests the conformity between the empirical distribution and the given theoretical distribution.

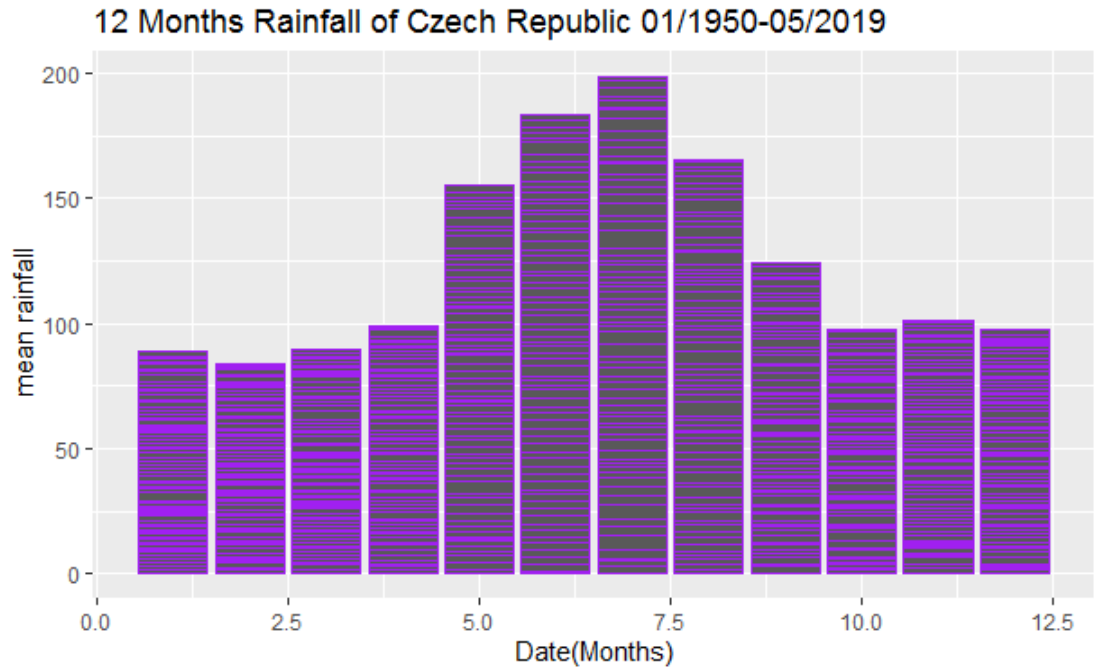
This below quantile figure shows that data is normally distributed, more than 50% medians lies on the line, it means that the points in the QQ-normal plot lie on a straight diagonal line and the deviations from the straight line are minimal.



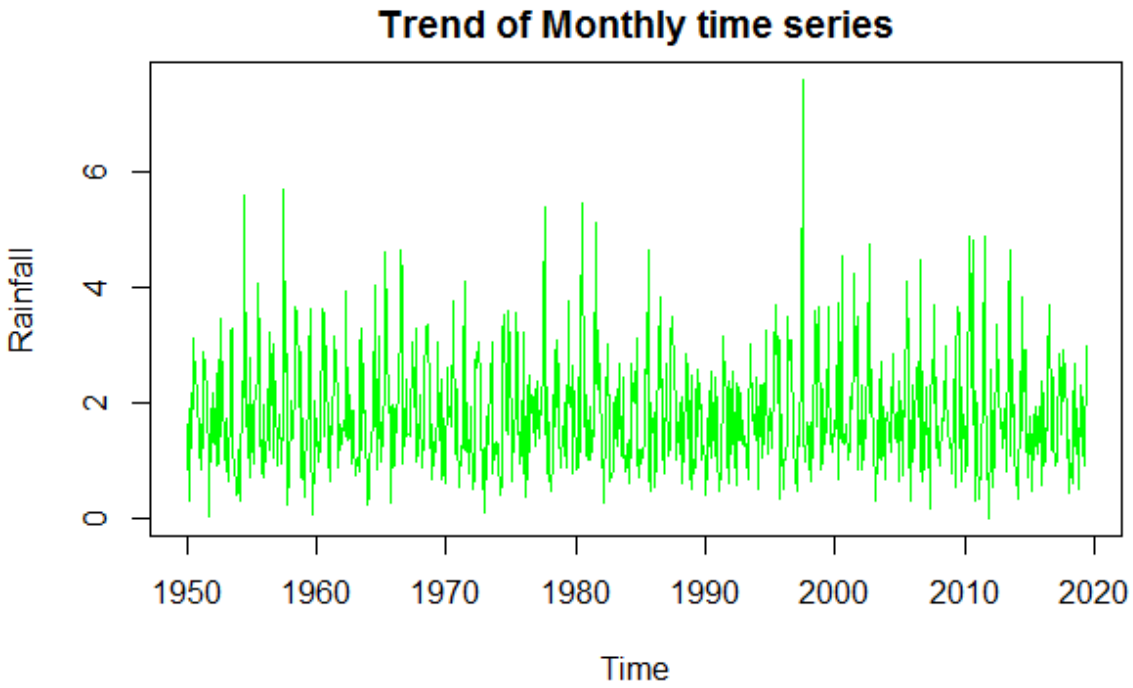


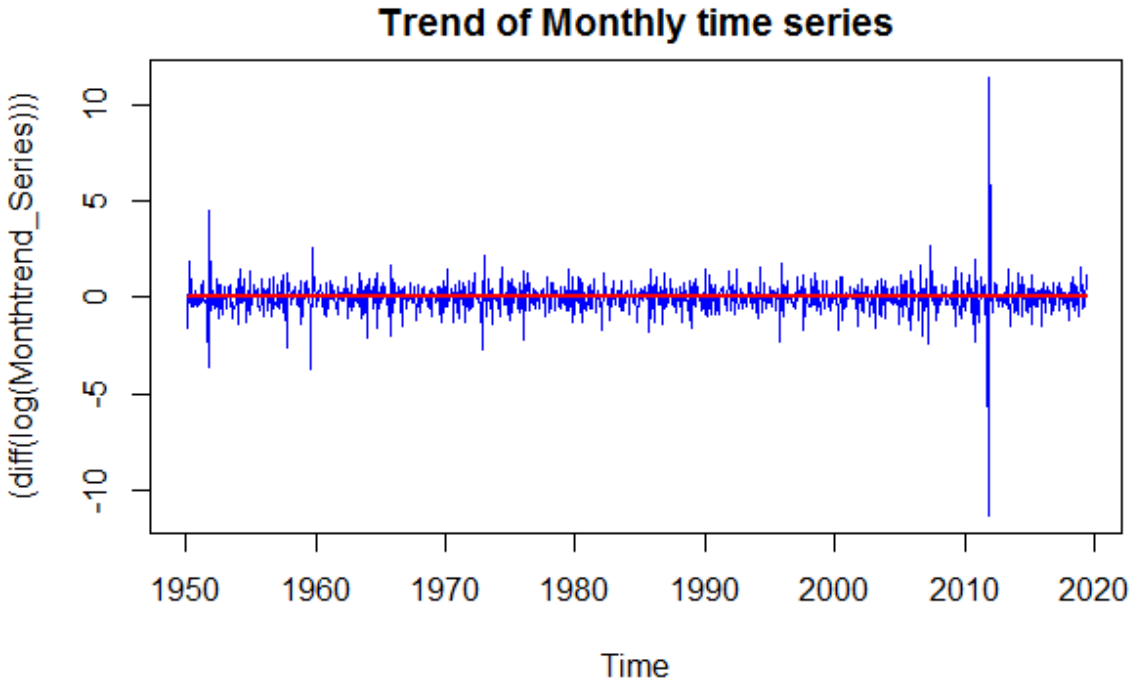
Boxplot consists of a "box" which goes from the first quartile (Q1) to the third quartile (Q3). Within the box, a vertical line is drawn at the Q2, the median of the data set which is mid-point of the data represents the middle 50% of scores for the group “Rainfall”. Two horizontal lines, called whiskers, extend from the front and back of the box.

Most of the data lies at the outlier box of month July and August which shows maximum value of rainfall and November outlier has lesser values than others.



To compare the monthly mean data set and track changes over time, a bar plot is used here best to explain the influence of precipitation of each month. We can see that the most frequent rainfall occurs at the month of June and July and lessor amount of rainfall falls in February.

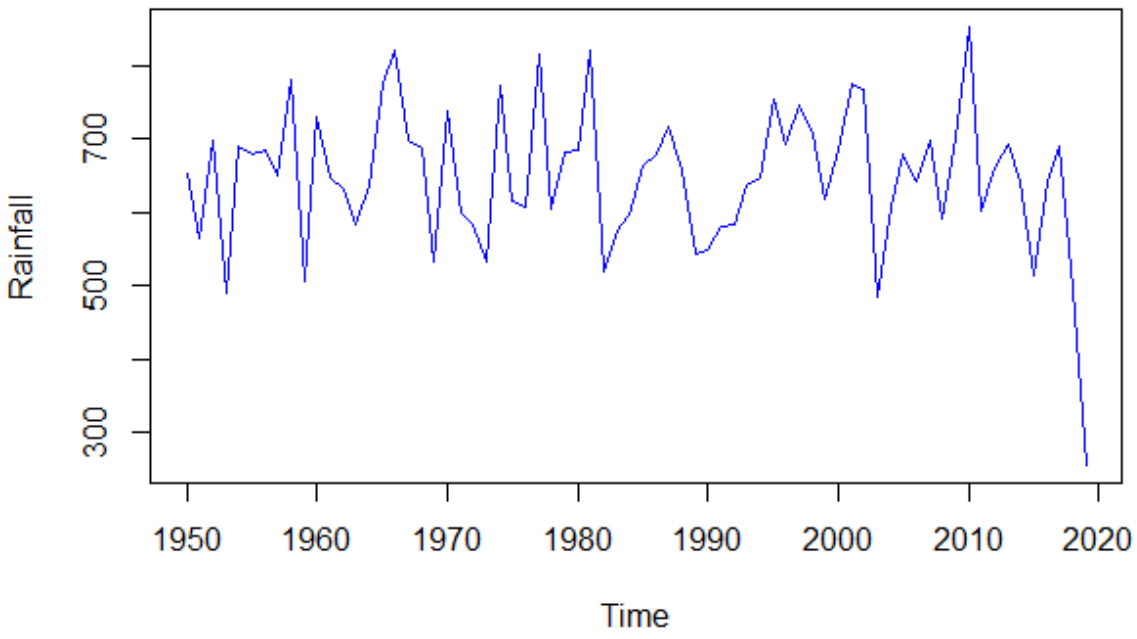


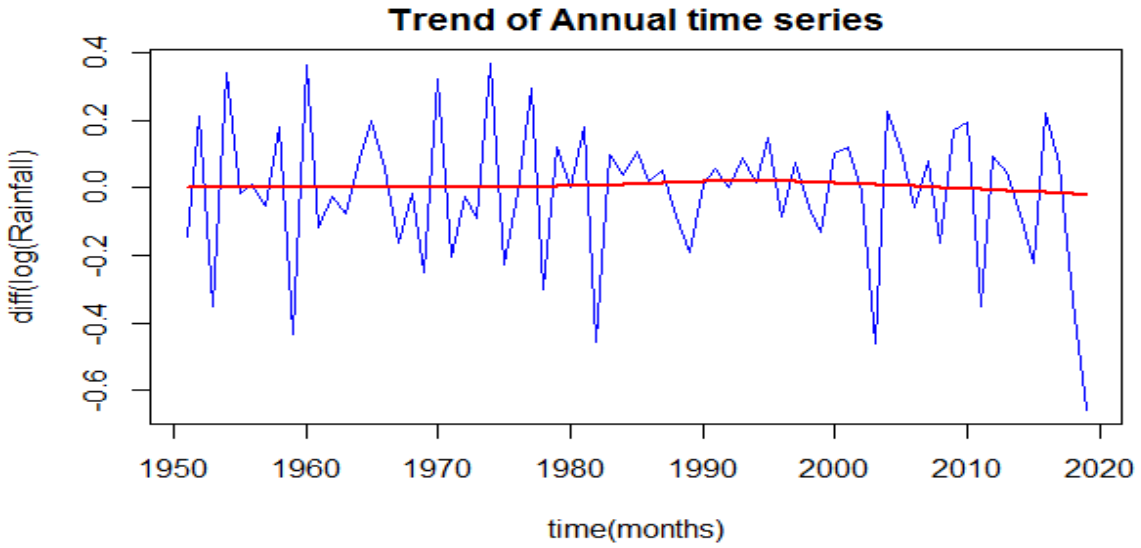


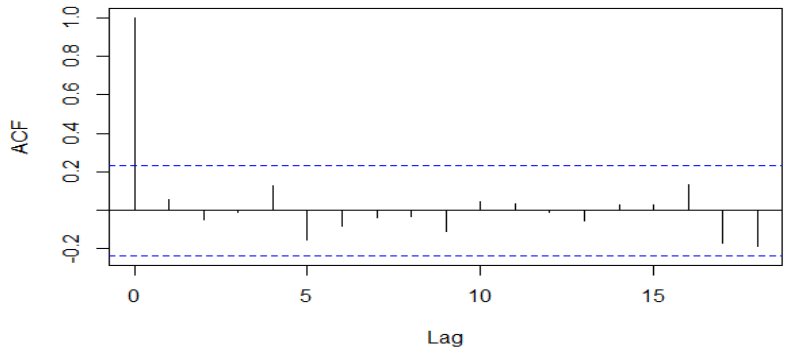
Monthly trend with line fitting does not produce clear picture either the trend is increasing or decreasing with time. The next task is to plot annual trend time series.

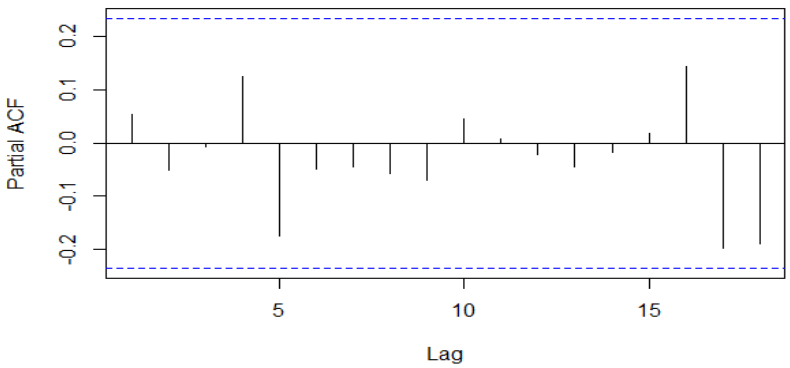
The time series plot produced by R reveals the presence of an downward trend in the annual precipitation levels for the entire Czech Republic over the period of interest.

To better see this trend, let us fit a nonparametric loess curve to the data using the lowess() function in R: For the annual result, the









The presence of serial correlation among the annual precipitation levels can be investigated visually with the help of the acf() and pacf() functions in R, which compute the autocorrelation and partial autocorrelation corresponding to the time series of annual precipitation levels.

Examining the autocorrelation and partial autocorrelation plots produced by R for the time series of annual precipitation levels suggests that the autocorrelation and partial autocorrelation present in this series do not appear significant. (Indeed, most of the vertical spikes in the ACF and Partial ACF plots produced by R fall within the horizontal band defined by the blue dotted lines beyond which autocorrelations and partial autocorrelations would be deemed to be significant.)
The output of this test produced by R is terse and reports just the tau value Kandall's tau statistic and the 2-sided p-value for testing the hypotheses "Ho: no trend" versus "Ha: monotonic trend (upward or downward)":

data: annual_trend_Series
tau = -0.0344, 2-sided pvalue =0.67762

To compute a 95% bootstrap confidence interval for the slope of the trend, use the boot.ci() function in the boot package Level Percentile
95% (-0.1685, 0.1695)

Conclusion : By looking our results, if we compare 70 years rainfall data of Czech Republic starts from January 1950 to May 2019. the trend value given us tau negative and p value <0.5 so it reveals that trend decreases with the rainfall decreases.