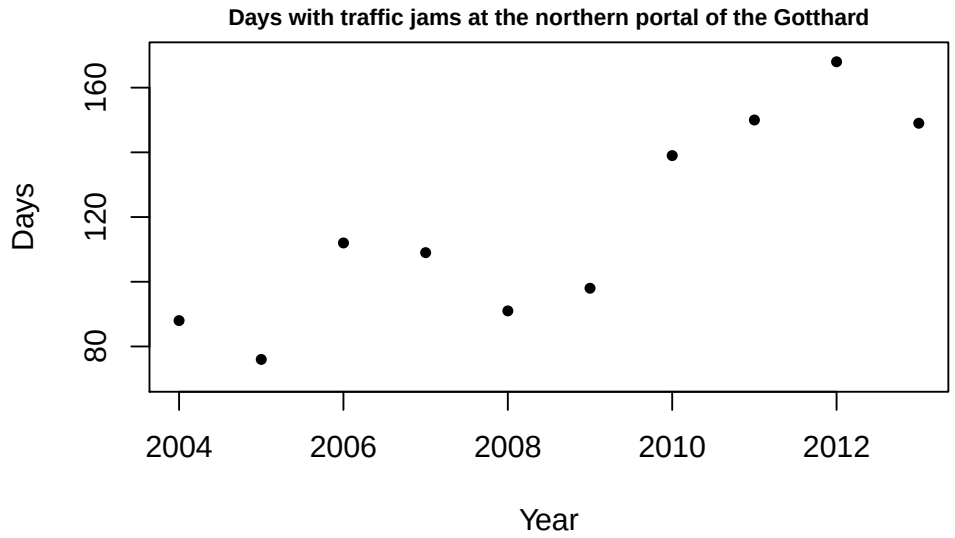
Series 2

1. In this exercise we look at a dataset on traffic jams at the northern portal of the Gotthard Tunnel. Specifically, the number of days with traffic jams are recorded for the years 2004 – 2013.

2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
88	76	112	109	91	98	139	150	168	149

This dataset is available from the website of ASTRA (Federal Roads Office).

Load the data with R and create a scatter plot. Since you will need to compute smoothers by hand in the following exercises, the scatter plot is reproduced here:



a) Compute a running mean smoother by hand, using a window width of three years and a step size of one year. Use the year 2004 for the first window center and use the year 2013 for the last one. Draw your solution into the provided scatter plot.

b) Now create the running mean smoother with R, using the function `ksmooth()`. Adjust the function call so that you reproduce the smoother computed by hand. Plot your solution together with a scatter plot and compare with your hand-drawn solution.

c) Create a Gaussian kernel smoother by hand. Choose the Gaussian distribution in such a way that the standard deviation is two years. Again, use the year 2004 for the first window center and use the year 2013 for the last one. Draw your solution into the provided scatter plot.

d) Now create the Gaussian kernel smoother with R, using the function `ksmooth()`, and add it to a scatter plot. Vary the choice of the smoothing parameter and choose an appropriate value. Plot the resulting smoother as well.

e) Finally, use a LOESS smoother for this data set and add the solution to the scatter plot. Vary the arguments `degree=`... and `span=`... and study the behavior of the fit.

2. In this problem we look at a dataset containing measurements of solar radiation hitting the earth. The data is stored in the file `solar.radiation.rda` which you can download from `https://stat.ethz.ch/education/semesters/as2015/asr/Uebungen/solar.radiation.rda`. The station is located in central Europe and the data set contains 29 observations, measured between 1963 and 2003. Note that for some years there are no measurements and some values might be corrupted!

a) Load the data in R and create a scatter plot. Add various smoothers to the scatter plot (Running Mean & Gaussian Kernel Smoother with bandwidth 10; LOESS).

b) Did the radiation intensity decrease over the years? Think carefully and explain your reasoning.

3. Which of the following statements are false and why?

a) In presence of outliers, in order to visualize the relationship between a response variable and a predictor variable it is better to use a Gaussian kernel smoother, rather than a LOESS smoother.

b) When using a smoother, it is easier to make predictions for values that lie outside of the range of the observed data, as compared to linear modeling methods.

c) A well-chosen smoother is fitted such that it minimizes the residual sum of squares, i.e. the distance between smoother and data points is minimized.

d) The following is an example of a linear model:
$y = \beta_0 + \beta_1 \sin(x_1) + \beta_2 \cos(x_1) + \beta_3 \log(x_2) + \beta_4 \sqrt{x_3} + E$

e) The following is an example of a linear model:
$y = \beta_1 x_1 + \exp(\beta_2 x_2) + E$

Preliminary discussion: Monday, September 21.

Deadline: Monday, September 28.