**The association of the cost of damage property caused by weather events with its characteristics.**   
  
**Introduction.**

This is my final project in the course [Data Analysis and Interpretation](https://www.coursera.org/specializations/data-analysis) organized by Wesleyan University (USA).  
The source of data is StormEvent Database from [the National Center for Environmental Information](href=%22https:/www.ncdc.noaa.gov/).  
The purpose of the research is to identify what characteristics of various weather events are associated with the cost of property destroyed by this particular event.   
The goal of the research to check the association of damage property with a climate region where it happened, month(relation to seasons), the scale of the event (did it happened on the county or zone level), the event duration and type of weather events.  
Although I am a biologist and not climatologist, it is important for me to gain an experience with data which is not related to biology or medicine.  
The result of the research could be used for minimizing the damage caused by weather events. In process of city planning, house building other kinds of similar decision making it is important to know where in which month and what kind of weather event can be associated with a serious damage.

**Methods**

**Sample**  
The dataset contains N = 166068 weather events that took a place between January 2013 and October 2015 in the United States of America. This dataset is a part of the official publication of the National Oceanic and Atmospheric Administration (NOAA). The part of information could have been provided not by the National Weather Service (NWS), but by the media, law enforcement and/or other government agencies, private companies, individuals etc. Beyond ordinary weather events, rare or unusual phenomena and some other meteorological events like maximum or minimum temperature were written. After the data management, only N = 37033 weather events were considered.

**Measures**  
As mentioned above, the response variable is damage caused by a weather event. The cost of damage was entered as actual dollar amounts, but only in case if reasonably accurate estimate could be found. The estimation was provided by an insurance company or other individuals who were qualified enough to perform the evaluation. Because not every event caused the damage and plenty of them were not correctly evaluated, data management was performed.

After first attempts of univariate analysis, it became obvious that

The potential predictors are:  
**1)** a month when the event happened;  
**2)** a region where the episode took a place;  
**3)** the type of the weather event (e.g. Flood, Marine Thunderstorm Wind etc).  
Probably, when I have some results, I will check some feature of the particular type of the weather event (like tornado type of magnitude).

**Analyses**  
Distribution of every predictor was evaluated by frequency table. The bar charts and box plots were examined. For a bivariate test of the association of a month, a region or type of weather event with the damage level, the Analysis of Variances (ANOVA) was used.  
To perform a deeper analysis, the decision tree and the random forest methods were implemented. For the machine learning approaches the dataset was divided into a training set (70%) N = 25923 and test set (30%) N = 11110.