# R Final Project: The Titanic

Daniel Alonso & Ander Iturburu

October 28th, 2020

## The Titanic disaster dataset

### Introduction

In this project we will explore the Titanic Passenger's dataset. We intend to make a surface description of the datasets through plots, analysis of such plots and detailed descriptions of our observations relating them. We will also use caret to analyze the relationship of one of our categorical variables in terms of other variables (in our case, mortality of the titanic's passengers).

The project will be divided in two parts: The descriptive analysis and the use of caret to relate the variables. Each part will explore a few key points that we consider quite interesting. We want to create a general idea of the content of the dataset in order to later on effectively know what might have affected the mortality of passengers in the titanic. The teachings from such a tragic event must've taught us something right?

A few key points to explore will be:

- How are our most relevant variables distributed?
- How do these variables respond to grouping by categorical variables in the dataset?
- Which variables are correlated to which?
- What is really relevant when it comes to determining a possible increase in risk of death of the passengers?

Among other things, which hopefully shall paint an accurate full picture when it comes to the risk of death in this tragic event.

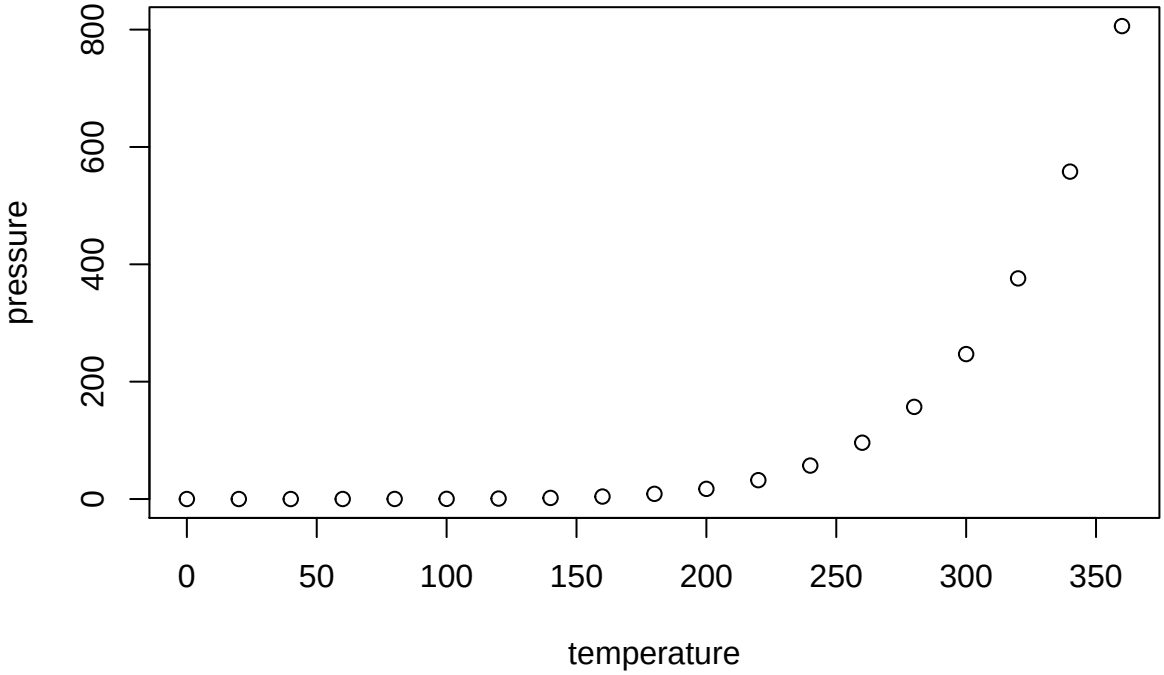
**Run this cell if you do NOT have the libraries we'll use for this project**

**First off let's load the libraries we'll use during this project**

```
library(dplyr)
library(ggplot2)
library(fitdistrplus)
library(PerformanceAnalytics)
library(reshape2)
library(vcd)
library(EnvStats)
library(scales)
```

## Descriptive analysis

### Our variables




Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.