

# Lab 1: Exploratory Analysis of CEO Salary Data

*Paul Laskowski*

*August 18, 2016*

## Introduction

You are hired by a think tank that is preparing a report on CEO salary. The think tank is especially interested in whether company performance is related to salary. They have collected data on a selection of companies, provided in the file `ceosal_w203.RData`. You are also given the following codebook:

1. salary	1900 compensation, \$1000s
2. age	in years
3. college	=1 if attended college
4. grad	=1 if attended graduate school
5. comten	years with company
6. ceoten	years as ceo with company
7. profits	1990 profits, millions
8. mktval	market value, end 1990, mills.

Note that this is a real data set, but it may have been modified by your instructors to test your abilities.

You are to conduct an exploratory analysis of this dataset to address the think tank's questions. Remember that you are to use descriptive tools, but note any features you find that you think would be relevant to statistical modeling.

## Deliverables

You must turn in

1. Your pdf report. In this report, do not suppress the R code that generates your output.
2. The R script you use to generate your report

For example, many students will find it convenient to write an Rmd file and turn it in along with the pdf it generates.

Be sure to follow the guidelines we covered in class. In particular, do not include any output that you do not discuss in your text.

Please include a brief introduction and a brief discussion that explains your findings at a high level.

Your analysis should be thorough, but limit your report to a maximum of 30 pages. This means that you will have to make choices about what variables and relationships to focus on (and justify those choices).

## Due Date

This lab is due 24 hours before the week 4 live session.