**Exploring the Evolution of Linux**

Find out about the evolution of the Linux operating system by exploring its version control system.

#### Project Description

Version control repositories like CVS, Subversion or Git store rich evolution information about a software project. In this project, you'll be challenged to read in, clean up and visualize a real world Git repository dataset of the Linux kernel. With almost 700k commits and thousands of contributors (find out the exact number in this project ;-) ) there are some little data cleaning and wrangling challenges that you'll encounter. But you'll also gain insights about the development activities over the last 13 years.

For this Project, you need to be familiar with Pandas DataFrames, especially the read\_csv and groupby functions, as well as with working with time series data. You can learn the required skills in these courses:

* [Intermediate Python for Data Science](https://www.datacamp.com/courses/intermediate-python-for-data-science)
* [pandas Foundations](https://www.datacamp.com/courses/pandas-foundations)

#### Project Tasks

* 1 Introduction
* 2 Reading in the dataset
* 3 Getting an overview
* 4 Finding the TOP 10 contributors
* 5 Wrangling the data
* 6 Treating wrong timestamps
* 7 Grouping commits per year
* 8 Visualizing the history of Linux
* 9 Conclusion