

# Group Project - Dataset

## AIDS Clinical Trial Group Study

The data for the group project are from a randomized, double-blind, study of AIDS patients with advanced immune suppression, that corresponds to CD4 counts of less than or equal to 50 cells/mm<sup>3</sup>.

The 1309 patients in this trial were randomized to one of four daily regimens of a medication called Zidovudine.

The categorical variable `treatment` in the dataset contains the treatment type information and is coded as follows:

Code	Treatment type
1	zidovudine alternating monthly with 400mg didanosine
2	zidovudine plus 2.25mg of zalcitabine
3	zidovudine plus 400mg of didanosine
4	zidovudine plus 400mg of didanosine plus 400mg of nevirapine

**Caution:** the `treatment` variable is a categorical variable, but it is coded as a numerical variable.

**Hint:** You need to change it to a `factor`.

There are other information gathered from the patients that are present in the dataset and are as follows:

Variable	Description
<code>week</code>	time since baseline (in weeks)
<code>age</code>	age (in years)
<code>gender</code>	male and female

Lastly, the response variable, `log_cd4`, is the log transformed CD4 counts; i.e.  $\log(\text{CD4 counts} + 1)$ , which is a **continuous** variable.

The **goal** of your project is to compare the effect of treatment types on the changes in log transformed CD4 counts over time.