# Exercise: Data Abstraction

Exercise Credit: [Enrico Bertini](http://enrico.bertini.io/teaching/), NYU Tandon

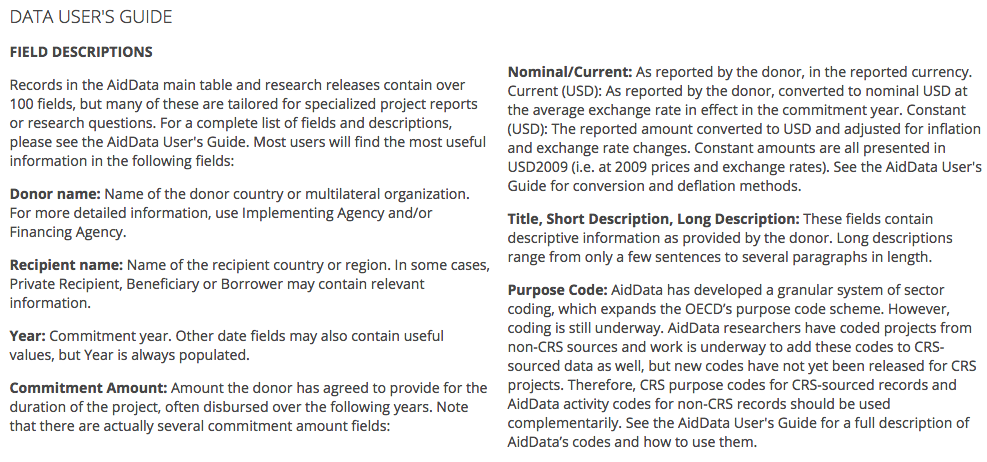
**Learning goals:**

* Learn how to recognize data set and attribute types.
* Learn how to generate data analysis questions and transform data in ways that enable you to answer them.

**How do you know if you are on the right path?** You can easily generate questions out of data, figure out how data needs to be transformed in order to answer the question and, identify what data set type and attribute types your data transformation has generated.

For this exercise we will use the “Aid Data” dataset. You can find a sample of the data in this week’s folder called “[Aid Data Example](https://docs.google.com/spreadsheets/d/1BsgFcb-4vgZv_Thh--RUrcInTfCJbtpkYkv0VypXXoQ/edit?usp=sharing)”. You can find more information on the data set here: <http://aiddata.org/user-guide>.

Here is an excerpt from the data set user’s guide that summarizes the meaning of the fields:



**EXERCISE:**

**Step 0: Make one googledoc for your group in Group Results folder**

* Document title should be “Group X” where X is your group number. Sweep from front to back, starting on the left: group 1 is the front left table, group 5 is back left, group 6 is front right, group 10 is back right.
* Put names of each group member at top of page.
* Copy what’s below, from Step 1 through the end, into the doc
* Work through these questions as a group.

**Step 1: Read the user guide and familiarize with the data set.**

* What is the meaning of each field?
* Is there any other important information you learned?
* Do you have any doubts/questions?

**Step 2: Perform an initial data abstraction step on the provided data set:**

* What is the data set type?
* What is the attribute type for each column/field?

**Step 3: Write 5 questions you would like to answer with this data set.**

**Step 4: Write for each question the following information:**

* Do you need a chart in order to answer this question?
* Which fields do you need to use to answer the question?
* Do you need to transform the data in order to answer the question? If yes, what transformations are needed?
* Do data set type and attribute type change when you need to transform the data? If yes, how do they change?

**REFLECT/DISCUSS:**

**What did you learn in this exercise?**

**How is this going to be useful in visualization design?**