

# CPSC 436V, Foundations 1

Out: Thu Jan 9 2020. Due: Wed Jan 15 2020, 6pm.

Submit through Canvas, as a file in either plain text (.txt) or PDF format.

## 0 Academic Integrity (0%, Mandatory)

Read this page and the two that it links to. Submitting this assignment is attesting that you fully understood that material.

<https://www.cs.ubc.ca/~tmm/courses/436V-20/cheat.html>

## 1 Data Abstraction: Attributes (5%)

What type of attribute (categorical, ordinal, quantitative) are the following?

1. 50 meter race times
2. College major
3. Amazon rating for a product
4. Product name
5. Weight of your favorite cat
6. Personality of your favorite cat

*Exercise Credit: adapted from [Lex & Meyer](#), Utah*

## 2 Data Abstraction: Central Park Squirrel Census (25%)

### [2018 Squirrel Census Data](#)

<https://data.cityofnewyork.us/Environment/2018-Central-Park-Squirrel-Census-Squirrel-Data/vfnx-vebw>

For each field

- Write down the attribute type (categorical, ordinal, quantitative)
- Determine its cardinality (number of levels) for categorical or ordinal, or range for quantitative

## 3 Data Abstraction: Foreign Aid (70%)

### [Foreign aid dataset](#)

For this dataset, read the field descriptions below, in domain language

**FIELD DESCRIPTIONS**

Records in the AidData main table and research releases contain over 100 fields, but many of these are tailored for specialized project reports or research questions. For a complete list of fields and descriptions, please see the AidData User's Guide. Most users will find the most useful information in the following fields:

**Donor name:** Name of the donor country or multilateral organization. For more detailed information, use Implementing Agency and/or Financing Agency.

**Recipient name:** Name of the recipient country or region. In some cases, Private Recipient, Beneficiary or Borrower may contain relevant information.

**Year:** Commitment year. Other date fields may also contain useful values, but Year is always populated.

**Commitment Amount:** Amount the donor has agreed to provide for the duration of the project, often disbursed over the following years. Note that there are actually several commitment amount fields:

**Nominal/Current:** As reported by the donor, in the reported currency. **Current (USD):** As reported by the donor, converted to nominal USD at the average exchange rate in effect in the commitment year. **Constant (USD):** The reported amount converted to USD and adjusted for inflation and exchange rate changes. Constant amounts are all presented in USD2009 (i.e. at 2009 prices and exchange rates). See the AidData User's Guide for conversion and deflation methods.

**Title, Short Description, Long Description:** These fields contain descriptive information as provided by the donor. Long descriptions range from only a few sentences to several paragraphs in length.

**Purpose Code:** AidData has developed a granular system of sector coding, which expands the OECD's purpose code scheme. However, coding is still underway. AidData researchers have coded projects from non-CRS sources and work is underway to add these codes to CRS-sourced data as well, but new codes have not yet been released for CRS projects. Therefore, CRS purpose codes for CRS-sourced records and AidData activity codes for non-CRS records should be used complementarily. See the AidData User's Guide for a full description of AidData's codes and how to use them.

**3.1: Overall**

- What is the dataset type(s)?
- How many fields/attributes does it have?
- How many items are there?

**3.2: Analyze for each field**

- Determine its cardinality/range.
- For categorical attributes, write down the number of unique levels
- For quantitative attributes, specify the range from min to max and note any other characterization that seems potentially useful (cyclic? Anything else?)
- For ordered attributes, consider whether it would be more useful to treat them categorical or quantitative, or to preserve them as ordered.

**3.3: Write 2 questions you would like to answer with this data set, from the point of view of an aid worker reporting to the government of a country providing aid.**

**3.4: For each question, write the following information:**

- Do you need a chart in order to answer this question?
- If none of your questions require a chart, try to create a few new ones that might benefit from one.
- Which fields/attributes 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?
- Do you have all the data you need to answer this question, or would you need additional data fields that are not provided here?

**3.5: Write 2 questions you would like to answer, from the point of view of a journalist reporting to the citizens of a country receiving aid.**

**3.6: For each new question, do step 3.4 above.**

