CODING ASSIGNMENT

To complete this assignment you will need to access these online resources:

www.opencontext.org

http://opencontext.org/projects/3

This assignment will guide you through the process of coding and analyzing data. It will also introduce you to open access resources in anthropology and how to make use of them.

Learning Outcomes

- Review quantitative and qualitative research methods, typology and coding
- Access open data and explore data management software
- Develop a classification and coding scheme
- Apply basic analytical techniques to describe the data

For this assignment, you will need access to a computer, with Internet (at least for Steps 1 and 2) and some form of data analysis software (such as Microsoft Excel or Apple Numbers).

If you aren't already familiar with your preferred data analysis software, it may be wise to take 10-15 minutes and do a quick online tutorial or watch a YouTube video for helpful tips.

Checklist to Submit

- ☐ Your finished dataset (pared down and coded)
- ☐ Research Report: A summary of your analysis (see template)

The Role of Code

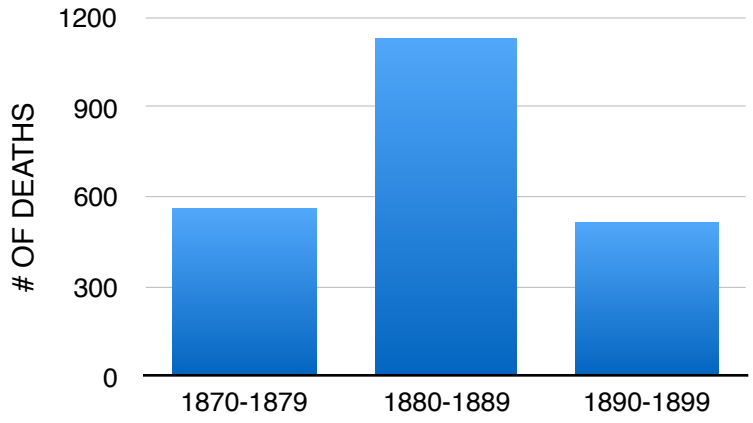
Anthropologists rely on coding to categorize both quantitative and qualitative data to facilitate further analysis. Coding makes data more readable for computer software, particularly for statistical analysis. If there are a number of researchers working on the same project, standardized codes can also minimize errors and increase reliability. However, for coded data to be useful, they must be consistent and comprehensive.

Coding Quantitative Data

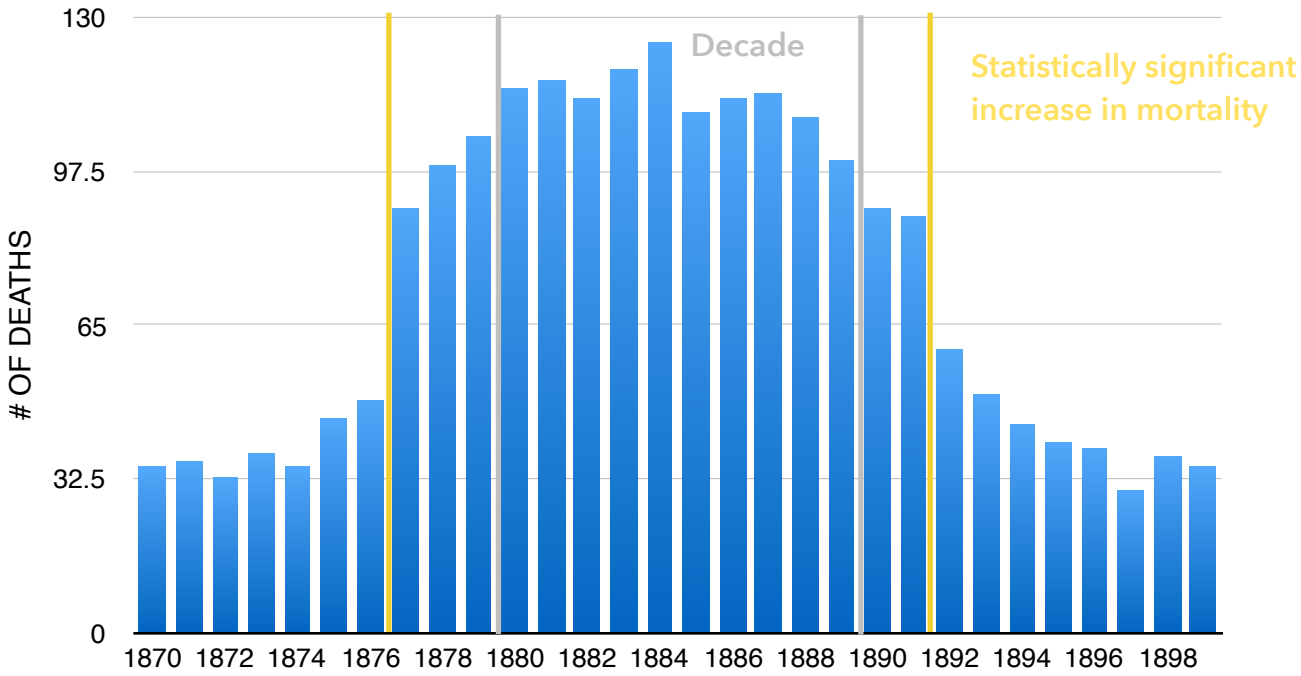
Although nominal or ordinal variables may already be ready to analyze, further coding may be used to create categories that are more relevant or succinct. For instance, dates (i.e. years) or ages of individuals are examples of quantitative data used in anthropology, however particularly when dealing with long time spans creating categories may simplify analysis. For example, ages may be categorized as follows:

Age	Category	Code
0-2.9	Infant	1
3-12	Child	2
13-19	Teenager	3
20-	Adult	4
–	No age 5	

Years could be categorized into decades or centuries, or other historic periods that are relevant to the research questions and variables. Although these categories may seem natural, remember that they are *always* artificial, imposed by the researcher(s). They will have a direct impact on the research results and may create artificial correlations so it is important to investigate data and think through your approach to analysis and thoroughly test results. For example, when breaking up dates into decade categories, you may separate the natural continuum of the data. Take for instance, the number of deaths recorded in a population. When broken up into decades, there is clearly a much higher mortality rate between 1880 and 1889.



You might also conclude that the average mortality rate for the population falls under 600 per decade, and that perhaps an epidemic, conflict or famine increased deaths during an isolated period of ten years.
However, if we look at the data for each year individually, the increase in mortality rates begins earlier and extends into the decade following:



If categories of nominal data end up masking the distribution of a particular variable, it may throw off interpretation and explanation.

Coding Qualitative Data

Anthropology relies on qualitative data for context. In ethnography, researchers may use codes to help identify trends and patterns in interviews or field notes. In these instances, researchers often begin by coding to uncover overall themes, before conducting a further round of more in-depth interpretive coding. In archaeology and biological anthropology, codes may be used to categorize a range of variables from site type or region when doing comparative analysis, to clustering artefact types (ceramics, lithics, faunal; personal adornment, domestic ware, weapon) or in biological anthropology, individual categories (biological sex) or pathologies. Prior to categorizing data, researchers will often review the data to familiarize themselves with variables. Their categories may also be informed by research conducted by other anthropologists, preliminary studies, and their research questions/hypotheses.

Coding can also help to streamline mixed methods approaches, making it easier and more direct to analyze the relationships between quantitative and qualitative research.

Open Access Anthropology

This assignment will make use of **open access** data: that is, online research data that are free of all restrictions that limit access or use (for instance, fees, copyright, licenses, etc.). The open movement, particularly in anthropology, is motivated by:

- the idea that all publicly funded research should be publicly available
- field and lab work is expensive and requires huge investments of time; with cuts to funding and other pressures, it is more sustainable to consider ways we can fully engage with existing data (which often remains underutilised)
- (social) science is based on the reproducibility and evaluation of results; publishing data means we can better evaluate the interpretations being presented
- no one can 'own' cultural knowledge, so licensing or selling of access is unethical
- traditional publication models favour large and wealthy businesses and institutions, but are unsustainable for researchers, communities and stakeholders
- the Internet provides the opportunity to truly embrace the openness of the web, to share data, interpretations and discussions in new ways

However, at the same time, open publication of data may not always be appropriate. For instance, ethnographic data may include confidential information about informants that may put them at physical or emotional risk. Releasing data about archaeological sites and artifacts may endanger them, increasing looting, destruction and theft. The politics of open data in anthropology should not be overlooked.

Open Context

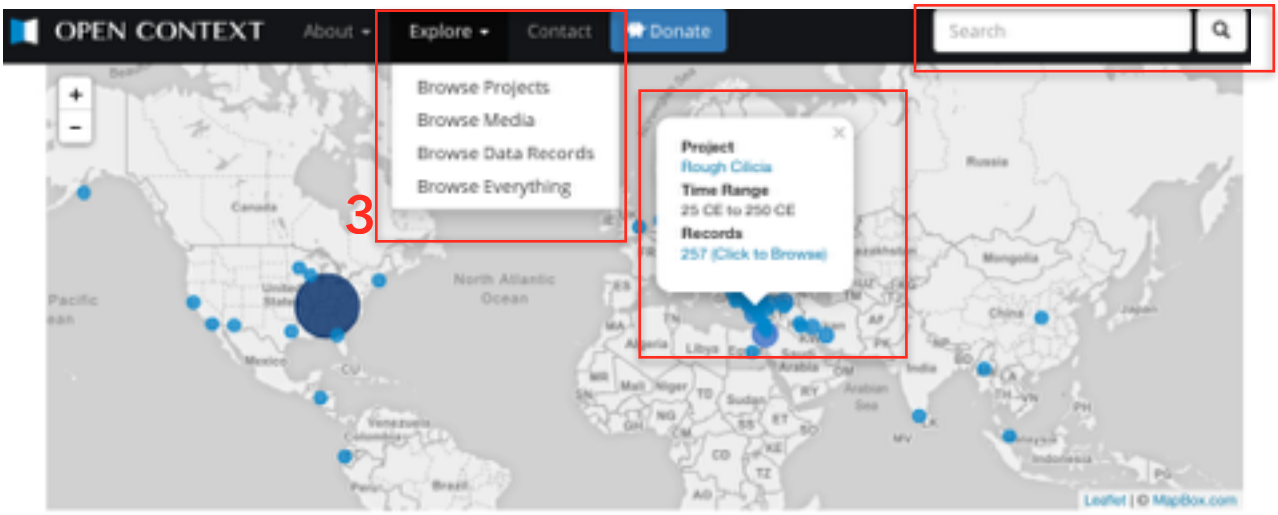
This assignment will utilize *Open Context*, a data publisher focussing on archaeology (with some biological archaeology). This not-for-profit organization was started by two archaeologists, Dr. Sarah Whitcher Kansa and Dr. Eric Kansa, recognizing the difficulties pursuing comparative research in archaeology due to a lack of shared data. This is not the only source of open access anthropology resources – there are online publishers of 3D models of hominid skeletons and ethnographic field notes, and other publishers of archaeological data. Open Context, however, have become leaders in open anthropology discourse and innovators in digital information management systems.

Take a moment to explore their website now: www.opencontext.org

STEP 1: Navigating Open Context:

From the home page, it is easy to start searching Open Context:

1) Visually, by clicking the blue dots on the map, which will bring up a brief description of project name, date and how many records are associated with it. You can zoom in and out of the map to see areas of interest/concentrations in more detail.
2) Textually, by entering search terms into the basic search function.
3) By 'theme' – choose from the Explore drop-down menu and see datasets listed by Project, Media, Data Records or Everything at once.



Projects: collection of data submitted to Open Context, often relating to a single site, but sometimes containing data related to a number of related sites. These will be accompanied with an abstract to help give you a sense of what was published and why.

Media: Type of media (images, video, etc.) associated with datasets; it may be relevant to certain research questions to only search datasets that have photographs submitted along with databases as additional evidence.

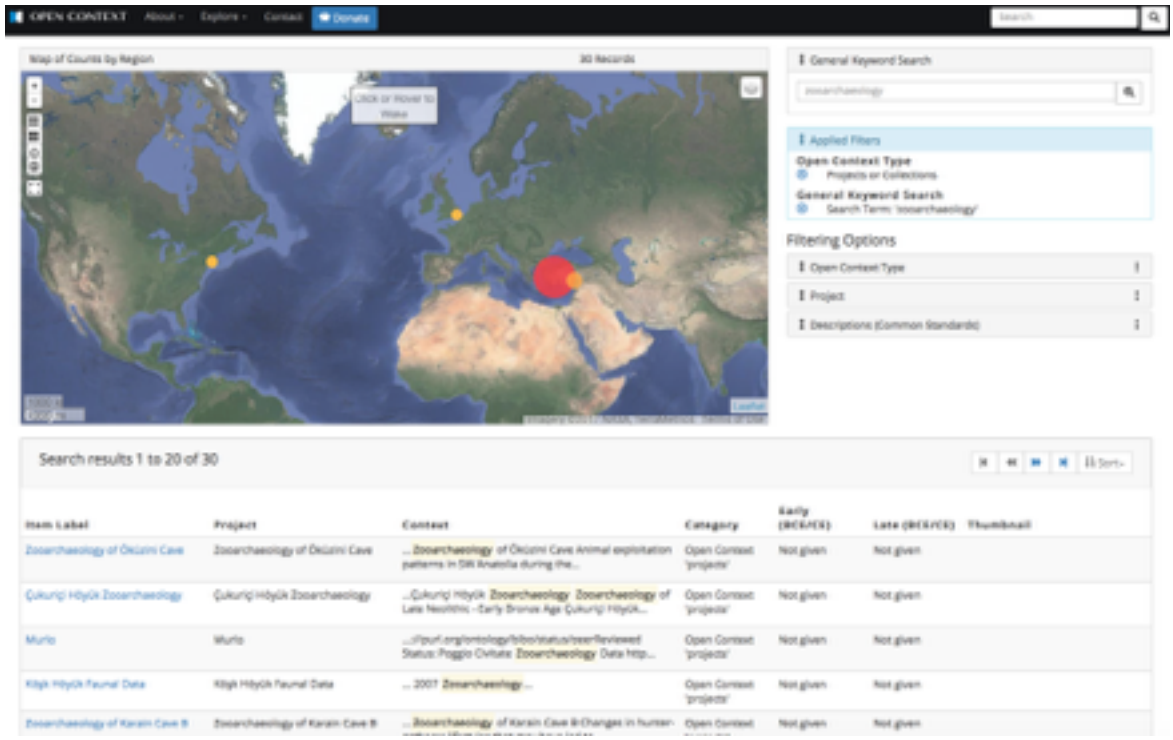
Data Records: Includes a range of types of 'items' or data, from field notes to raw data published in .csv files (which can easily be downloaded and imported into analytical software)

<u>Try typing 'zooarchaeology' into the search box.</u> You may notice that the search returns thousands of results (25,522 as of August 2017!) It would take far too much time to go through every single item returned in this search, let's narrow it further.

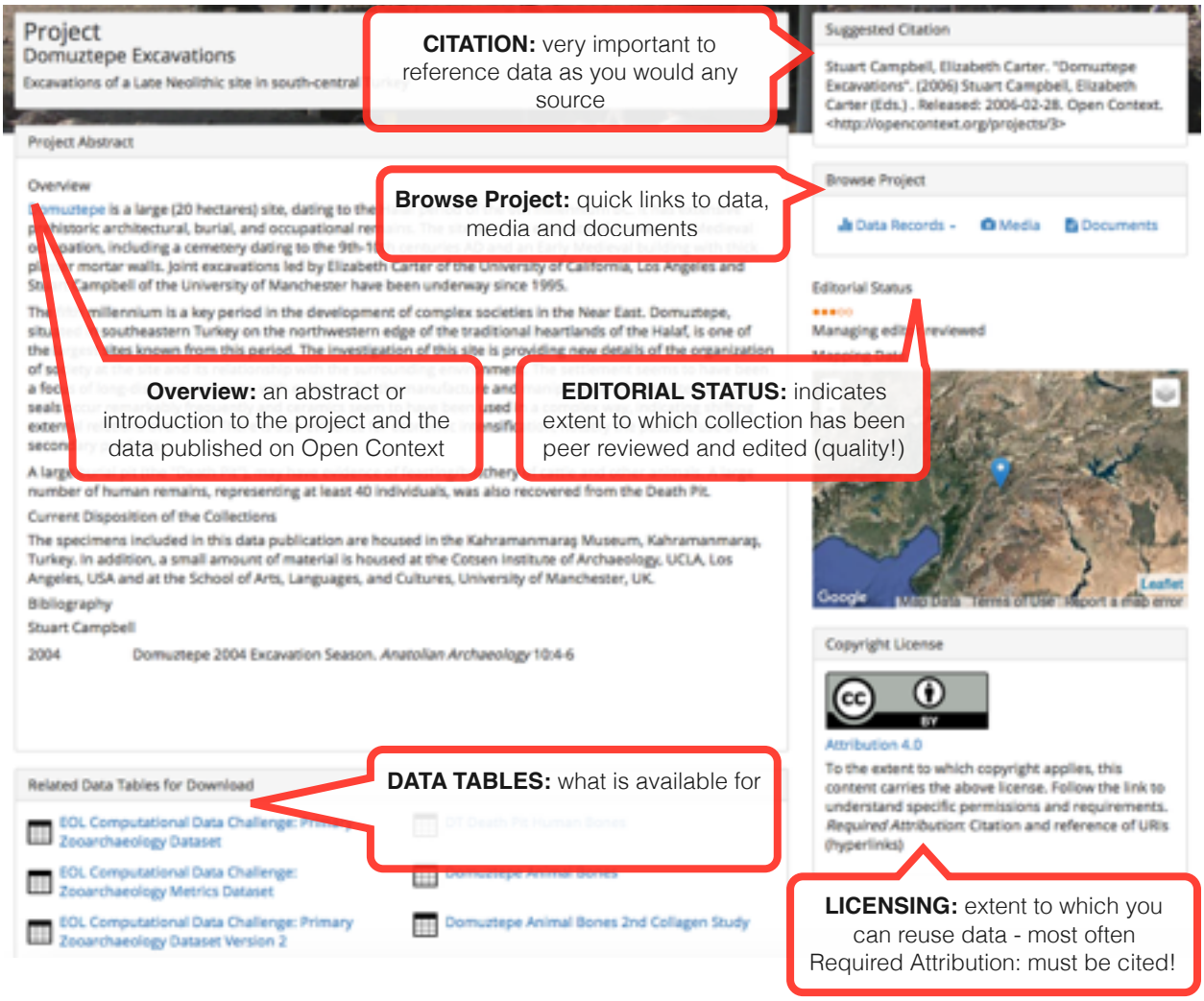
We can again click on the map to visually/geographically identify zoo archaeology records, or we can use the filters on the right side of the page to concentrate on relevant results.

Filter options include filtering by the 'Context' (geographic) or pinpointing a particular 'Project'. 'Open Context type' allows you to filter results by all *Subjects* (individual document or data item results, may include metadata, media, or raw data tables), *Projects or Collections*, which will just return larger groupings of results, usually by site, *Tables*, which will return raw datasets (.csv files), *Documents*, or *Descriptive* types. Projects or Collections and Tables are often the fastest way to filter the results into a more digestible format.

<u>Because we are interested in finding raw data with which we can work, filter your results by 'Projects' now.</u>



You should now have only around 30 or so results (as of August 2017). This is much more manageable! They are listed by Projects (which are hyperlinked), but have brief details on context, category and dates. **Go ahead and explore a project or two now** – what kind of information and downloads do they have associated with them?



Now that you can find data, how can you tell what is GOOD data to reuse in your research?

- **Editorial Status:** data goes through stages of editing (where someone will review consistency and organization of data) and peer review (other specialists in the field will review the quality, accuracy and precision of the data, as a peer reviewer would judge the quality and relevance of an article for publication) – while this isn't a guarantee, choosing data that has been through editing and/or peer review will often help to ensure good data

- **Metadata:** Look at the media and documents associated with the record - are they easy to navigate? consistent? are photographs good quality image files? These are often indications of not only the consistency and completeness of all records – if the documents are in disarray, often the data will be too! At the same time, if there is a lot of contextual data associated with records, you will have lots of information to help make sense of the data!
- **Download a Data Table** (or Two): If you want to know the quality of the data, take a look! Do you immediately notice spelling mistakes? or weird categories that you don't understand? How many items (artifacts, etc.) are included? What variables are there? Will this be enough of the right kinds of data to help you research the questions you are interested in?

Remember – look for edited data, lots of metadata (accompaniments) and clean, consistent data tables.

STEP 2: Downloading & Thinking About Data

If you haven't done so already, **navigate to the Domuztepe Excavations Project** on Open Context:

> Stuart Campbell, Elizabeth Carter. "Domuztepe Excavations". (2006) Stuart Campbell, Elizabeth Carter (Eds.) . Released: 2006-02-28. Open Context. <http://opencontext.org/projects/3>

We will use this project for the rest of this assignment. Read the Project Abstract and explore the items published with this project. **What types of research questions can you answer about Domuztepe using zooarchaeological remains?** Review the summary of the Domuztepe excavation if necessary: https://opencontext.org/projects/3.

The Data Tables for Download are hyperlinked to a Table Download page. **Download and open the data table file for 'Domuztepe Animal Bones'.**

Click the Download CSV file in the right corner. If it doesn't work try right clicking on the record – which should give you the option to 'Download linked file as…' and prompt you to choose where to save the file and what to name it. If it doesn't automatically open on your computer, go to the folder where it downloaded to and open it manually.

These data tables should be .csv files – CSV or "comma-separated values" is a very simple file format used to store data tables. Because this is one of the most simple file types, it is compatible with most analytical software, such as Microsoft Excel or Apple Numbers. However, this is essentially the text file version of data – if you make changes to formatting (such as the width of columns or making the text bold or a particular colour), none of this will save. It is best to immediately go to SAVE AS and save the file in the format of your software (i.e. an Excel file) so that your formatting will stick.

This might be a good time to make the file a bit easier to work with. Try making the header row 'bold' and freeze the frame so that it is always visible.

Take a look - how many variables are there (the top row – including everything from geographic information to bioarchaeological data)? What is the total size of the sample (i.e. how many bones)?

There are _____ variable columns and _____ rows.

That is a lot of data! We'll start by making it more manageable!

STEP 3: Pare down the Data

This dataset has been put together with the archaeological principle of preserving as much information as possible to be able to answer as many questions that future anthropologists may have. However, this means that there are a lot of variables that we don't need for this assignment.

Start by selecting and deleting the columns you feel you don't need. Be sure to KEEP:

Context 3

Taxon

Element

Side

Fusion (Proximal)

Sex

Six columns is already a lot easier to deal with - but we can still pare down further!

Sampling

Researchers will often create 'samples' — sometimes a sample will be inherent in your data. For instance, maybe you could only get in touch with ten out of fifty community members to interview, or maybe a lot of bones were not recovered from a site due to looting or development. In other cases, researchers will actively sample from data - selecting only the artifacts recovered from two middens on a site rather than the whole site, with the intention of comparing and contrasting those two contexts.

Regardless of the reasons or issues that contributed to your sample, it is important to recognize that sampling does have an impact on your results, and this impact should be considered while interpreting the data and presenting the results to other people.

To reduce the size of our dataset and make it more manageable for this assignment, we are going to use a **random sample**. A random sample is a way of creating a smaller subset of data without allowing your own subjectivities to colour your sample (for instance, picking out the 'best' candidates for your analysis may skew the results to your own assumptions or preconceptions).

Download YOUR random sample from CourseSpaces. This dataset has already been pared down to a random sample of bones (n=100) for you to work with.

However, if in future you need to create your own random sample, you simply have to follow these steps in whatever analytical software you are using:

1. Right-click the far left column's name. ...
2. Click Insert. ...
3. Select the new "A1" cell.
4. Type "= RAND()" into this cell. ...
5. Press ↵ Enter . ...
6. Select the cell with the random sample number.
7. Hold down the corner of the cell and drag to fill the whole column.
8. Select all the data and sort by the Random sample column (A1).
9. Select all data you don't need and delete it! (for instance, for a sample size of 100, select all data from 101 onwards)

Remember as you move forwards that this is a much smaller sample than you started with – you can calculate the proportion that your sample represents by dividing the sample by the original dataset size, and multiply by 100.

$$100/9265 \times 100 = 1.07\%$$

Even though it will take time to code and analyze 100 bones, this only represents 1.07% of the original sample! **What are your thoughts, are the results of your analysis going to be representative?**

STEP 4: Code the Data

If necessary, return to reading assignment on Undocumented Migration and review **typologies** before moving to this section of the assignment.

When using open data, other researchers have already started to create typologies, but it doesn't mean that all the work is done!

Start by creating a list of all the categories currently listed in Taxon, Element, Side, Fusion and Sex columns. Are there any categories that could be lumped together? For instance, unidentified and unknown are the same analytically. You may also notice *Ovis aries* (sheep) and *Capra hircus* (goat) are often lumped together because they are difficult to differentiate. Some researchers create an entire 'Sheep/Goat' category to include all Sheep, Goat, and uncertain Sheep/Goat identifications to end up with a larger sample and mitigate issues with uncertain identifications, while others like to keep certain identifications separate, though this may obscure patterns by limiting sample sizes. It is up to you which route to take – think about the size of the sample you have to work with and the impact it will have on analysis.

Once you have finalized your categories, coding can help streamline analysis and remove any spelling variations/typos that will make analysis more complicated. For instance within your species category, you may choose to code Sheep as 1, Goat as 2 and Sheep/Goat as 3. **Create a Key that "translates" all your codes so you can keep track.**

STEP 5: Basic Analysis

Archaeologists and biological anthropologists (among others) use **Minimum Number of Individuals (MNI)** to refer estimate the number of animals or people represented by a particular site or context. It is based on analysis of the least number of individuals that could possibly be present through logical inductive reasoning. It differs from **NISP (Number of Identified Specimens or Number of Individual Specimens)** - which is a raw count of how many bones or fragments of bones are present.

If there are three skulls amongst the remains, it is easy to reason that there are at least three individuals. It gets more complicated with skeletal elements that have right and left sides. If there are five femurs, and three are left femurs while two are right femurs, the MNI is three – as the two right femurs could potentially belong to the same individual as the left. It gets further broken down by age and sex, where possible. Fragmented bones further complicate this process – if there are two fragments of a left femur, but one is the distal (furthest from the centre of the body) end and one is proximal (closest to the centre of the body). The minimum number does, however, have the tendency to underestimate the number of animals actually represented while NISP may overestimate them; this must be taken into account when interpreting the data.

Start by calculating **NISP** and **MNI** for each species or taxon category in your sample. Use the table below as a starting point.

Taxon	NISP	MNI
Ovis aries / Capra hircus		
Sus scrofa		
Bos taurus		
Mollusca		

What does NISP and MNI tell us about Domuztepe?

What other analysis can you do with the variables in the dataset (context, NISP, taxon, element, side, fusion and sex?)

Research Report (Template of What to Submit)

STUDENT NAME: _____

SAMPLE DATASET #: _____

Question/Problem: [2-3 sentences]

1. What types of research questions can you answer about Domuztepe using zooarchaeological remains, particularly species, NISP and MNI?

Methods: [short paragraph]

1. Describe your process of sampling and coding. Be sure to answer:What was the original sample size and how many variables? What was your sample size and how many variables? Include a coding key and explanation of your choices.
2. Are the results of your analysis going to be representative? Why are they relevant?

Analysis: [1-2 paragraphs]

1. Submit any tables or figures created in your research process, i.e.:

Taxon	NISP	MNI
Ovis aries / Capra hircus		
Sus scrofa		
Bos taurus		
Mollusca		

2. Describe the results — any important observations about NISP, MNI, species distribution, etc.?

Conclusions: [short paragraph]

1. What does NISP and MNI tell us about Domuztepe?
2. What other analysis can you do with the variables in the dataset (context, NISP, taxon, element, side, fusion and sex?)

References Cited:

1. Be sure to reference the data that you use from Open Context! (see example on page 8)
2. Reference any additional literature that you use in relation to this project.

NOTES ON FORMATTING:

1. Make sure you indicate on your assignment which sample dataset you were assigned.
2. When we talk about data in research, it is important to give indications of sample and scale. For instance, if you write “50% of the remains were *Bos taurus*”, it makes a big difference if you are referring to 2 bones in a sample of 4, vs. 500 bones in a sample of 1000. The common way to insert this into scientific research reports is to use n=#/total sample, i.e. “50% of the remains were *Bos taurus* (n=500/1000).” This way, readers can contextualise the significance of your results and the confidence that we should have in them.
3. When using binomial nomenclature, referring to animals with genus and species names, be sure to italicize them, i.e. *Bos taurus*, *Ovis aries*.
4. Research reports are often written in a formal style; avoid using colloquial language structures (including contractions) or first person references, and be sure to always reference any sources of data, ideas or literature (trying to rely solely on academic sources)