1. Administrative questions
2. Today
   1. In our pyramid of sources, processed and presented, we’re still really looking at the stuff. We’ve moved from the way we decide which stuff fits in our project to the way we divide that stuff up into chunks we call data.
   2. I’ll argue that what we call “data” is not objective, but constructed through interpretation.
   3. Then I’ll explain some different ways of organizing and sorting out data
   4. My goal is for you to understand that even though databases often seem neutral or scientific, in reality, they are rhetorical devices that shape the kinds of knowledge we’re able to obtain from them.
3. The concept of data
   1. How do you define data?
   2. The word data comes from the Latin “to give,” which makes it seem as though it’s a given — that it’s something naturally obtained from the world.
   3. But I want to argue today that data itself is subjective, in much the same way we agreed that classification is subjective.
   4. There is no natural or necessary relationship between the world as it exists and the data we can gather about it.
   5. The dictionary definition of data: “quantities, characters, or symbols on which operations are performed by a computer, considered collectively.”
   6. So in order to create data, we have to somehow divide stuff up into parts — we have to decide how big or small those gridlines are going to be. And in order to do *that*, we have to come up with some units of measurement.
   7. What are some things you could measure about this classroom? What units of measurements might you use?
4. Parameterization
   1. It’s interesting to learn that even units of measurement have very specific histories.
   2. The process of coming up with these units is known as parameterization, and the measurements themselves are called parameters.
   3. Parameters are artificial. And they often differ among different groups.
   4. Farenheit: invented by Daniel Gabriel Greenberg in 1724, based on the freezing point of water, the boiling point of water, and human body temperature. So very human-centric.
   5. Time is another measurement that varies from place to place.
      1. Classical China: decimal time, day divided into 100 hours.
      2. Hindu chronology: day divided into three parts.
      3. Even our calendar, which has 365 days, doesn’t always work out — which is why we have leap years.
   6. In Scotland and England, people made use of the herring barrel, which measured 1000 herrings.
   7. A fathom is a length of rope that can be held outstretched between two arms: six feet.
   8. Even numbers don’t have the same meaning from place to place.
      1. British billion = a million million
      2. American billion = a thousand million
   9. More dramatically: the 3/5ths compromise was a way of measuring people.
      1. 1787, compromise between Southern and Northern states in which 3/5ths of the population of slaves would be counted for the purposes of taxation.
   10. Can you imagine a way of measuring something meaningful about your life? What would your unit of measurement be?
   11. Just as we know classification is subjective, we know that parameterization is subjective, too.
5. Data and capta
   1. Just as we have to accept the constructedness of classification and then move on, we have to accept the constructedness of data if we’re going to use computers or operate in modern life.
   2. In this class, we suggest a new name for data: capta!
   3. We’re so accustomed to associating data with what is natural and given that it’s hard to see how artificial it is.
   4. But the word capta means created or constructed — so let’s think of data as capta!
6. From data to databases
   1. But the story gets even more complicated, because we don’t just divide reality up into data.
   2. We also need to store data and connect it together.
   3. In order to store data and associate one piece of data with another, we create databases.
   4. There are many different kinds of databases: we’ll talk about two, the flat database and the relational database.
   5. Does anyone recognize this?
   6. It’s a spreadsheet, or what is also known as a flat database.
   7. The spreadsheet is pretty ubiquitous in our historical moment. But just like units of measurement have a history, this method of organizing data also has a history.
7. Descended from an accounting book, designed to keep track of profit and loss
   1. Here is an example from 1905
8. VisiCalc
   1. Invented 1979, sold over 100,000 copies
   2. Also specifically designed for accounting
   3. So important to remember that when we’re entering data into a spreadsheet, we’re adapting our information into a system that was originally designed for accounting.
9. Database vocabulary
   1. A database is only as good as the thought you put into constructing it.
   2. Easy to construct a database and find out later that it doesn’t do what you wanted it to do.
   3. Content model
   4. Content type
   5. Record
   6. Table
10. Star Paws
    1. Imagine that your uncle owned a pet talent agency. When he died, he left you with all of his records, which he stored on index cards.
    2. As a team, create records for ten pets, including the pet’s name, its talent, the roles it played, the pet description, and any other relevant information.
    3. Now, figure out what the content types will be and create a spreadsheet.
    4. Enter a few pets into your spreadsheet.
    5. Do you notice any differences in the way you’ve organized or entered information? Can you find all the dogs with a particular talent? Can you find all the dogs with a particular talent that live in Santa Monica?
11. Spreadsheets to Relational Databases I
    1. But spreadsheets have some limitations. What if one owner has multiple pets?
    2. Do you have to enter the same information multiple times?
    3. What is the owner’s phone number changes?
    4. Do you have to find every instance of the owner to change the phone number?
    5. Once you’re collecting data about a few different types of things, you may run into trouble if you’re using a flat database.
12. Spreadsheets to Relational Databases II
    1. This problem led to the development of relational databases in the 1960s.
    2. In a relational database, every kind of thing gets its own table.
    3. So “pets” is one table and “owners” is another kind of table.
    4. The tables are linked together using numerical keys.
    5. Paragraph analogy.
    6. Each table should have one theme.
    7. Separating the tables allows you to change information in one place and ensure that information remains consistent throughout the database.
    8. Spreadsheet software: Excel, Google Spreadsheet. Relational database: Access, Filemaker, Bento.
    9. The relationship I’ve drawn here is called one to many, because each pet can only have one owner, but each owner can have many pets.
    10. You can diagram multiple relationships with pen and paper. Try doing that with pets, owners, clients, and roles. Are these relationships one to many? Many to one?
13. We call this diagram a data model, and it can range from fairly simple to incredibly complicated.
14. Database and rhetoric
    1. If a database table is like a paragraph, then the way you connect each table together is like the rhetorical mode of an essay: it’s the way you connect themes together in order to develop meaning.
    2. If two tables aren’t connected, you can’t discern a relationship.
15. Kindred Britain
    1. It, too, has a data model.
16. Conclusion
    1. So we see that not only is data a construct — data is really capta — but the way we link it together is also a product of decisions that affect what we can learn from the world.