Database = relational database

* Spreadsheet is not a database – database assumes
* Database = a collection of tables with data, these tables have columns and rows
* Humans gives instructions to database manager to make tables of the data, by the questions we ask it
* Database manager is the intermediate between the database and us (it means we are not messing up data every time we want to ‘see’ something) – SQL lite is our database manager – run query against data to make a databased table

Database full of tables

* Row = value
* Column = attribute of this value
* Actions you make in a Database can move around rows – this cannot be done in a spreadsheet

SQL

* Structured query language used in databases
* All sequel commands are terminated by a semi-colon

Styles for SQL

* Commands are all *CAPS* (BIRD)
* Columns (fields/attributes) are in all *lowercase* (bird)
* Tables are first letter Caps (Bird)

Primary key

* Want to have specific Identifiers (primary keys [the ID] is the column or columns which intersect your row)

Commanding and querying

* You can ask for only data you want
* The order you request in the command is how it will structure the table
* Add ‘--' make this before something to make a comment

Key words

* Order of keywords is very strict
* SELECT

[SELECT DISTINCT] – summarises data, shows you each *KIND* of row *ONCE*SELECT min(column) will give the lowest amount

* FROM
* WHERE – make a conditional (e.g. WHERE ‘column 2’ = ‘this thing’) – limits which rows are returned based on condition (‘and’ and ‘or’ can be combined)  
  WHERE can take ‘IN’ (thing 1, thing 2) to get number of columns/rows.   
  WHERE can take ‘LIKE’ – Like is the ability to find something within a whole word. E.g. one could search any verbs in Latin of *uoco* within a text by searching ‘uoc%’.
* GROUP BY -
* ORDER BY – orders data (takes columns – automatically take alphabetically) –   
  ‘DESC’ after first column input = descending   
  ‘ASC’ for ascending, but unnecessary [‘random()’ is a *function* and will put them in random order]
* LIMIT –
* AS = rename the output column
* *IS NULL* – is how to find *null*
* *Aggregate group functions*  
  Avg() gives average, count() gives count, min() gives min, max() gives max

Blanks (*null*)

* There is difference between this is left blank, and this value is blank
* *True*
* *False*
* *Null* – value has not been determined (absence of entry)

*Null* behave weirdly   
🍆 in my system

* IS NULL – is how to find *null* (and IS NOT NULL) if you want everything else.

Joins

* How to combine multiple tables – JOIN
* One must join them on ‘the key’
* Based upon PRIMARY KEYS and their links with FOREIGN KEYS in different tables (FK= columns with a value used in a Primary Key but in a different table)

Theory:

* **Atomicity**
* Each field (attribute) should contain one observation
* Each field (attribute/column) should contain something that is NOT split-able (e.g. first name and last name, NOT full name)
* **Key**
* Each table should have a Primary key
* A Primary key should be one attribute
* If there is no natural ‘key’, make the primary key a number
* **Redundancy**
* There must be no redundant information
* To avoid anomalies – each observation of ‘reality’ (data) must be singular
* When you cannot distinguish which data is not correct
* Any piece of data should occur once and only once in your database
* **Units**
* Every value should have a unit
* Store in column
* No ambiguous units

**DataMaker**

* CREATE TABLE is a command to create a table
* DROP TABLE is a command to expunge the table
* ALWAYS keep an .sql text file back up

How to input data line by line:

* INSERT INTO  
    
  How to delete data
* DELETE  
  FROM

How to edit data

* UPDATE “table”
* SET “column” = 'Shanks'
* WHERE “primary key” = “row”

**For Brian if we meet**

Come with

* Data sample
* What you want to ask of the data
* What fields are multifaceted