Week 4 notes

**Spreadsheet basics (Google sheets or Microsoft Excel)**

Attribute – column names (a characteristic), column labels, headers, header row

Observation – a row, we see all the attributes (think of it as one data point)

Formula – a set of instructions that perform a specific action

**To sort data in excel by certain column names:**

Highlight the entire dataset, then go to the tab: Data 🡪 Sort

Then select the column name you want to sort by

Google sheets training

[Google Sheets training and help - Google Workspace Learning Center](https://support.google.com/a/users/answer/9282959?visit_id=637361702049227170-1815413770&rd=1)

Google Sheets cheat sheet

[Google Sheets cheat sheet - Google Workspace Learning Center](https://support.google.com/a/users/answer/9300022)

Microsoft Excel Training

[Excel video training (microsoft.com)](https://support.microsoft.com/en-us/office/excel-video-training-9bc05390-e94c-46af-a5b3-d7c22f6990bb)

**SQL – a sneak peak**

Queries are universal using SQL – whatever database you use, you can use SQL to query everything

Note:

using the asterisk (\*) means selecting all i.e

SELECT \*

Basic structure:

SELECT

Columns you want to look at (can have multiple eg. First\_name, last\_name, id)

FROM

Table the data lives in

WHERE

Certain condition/conditions is/are met (AND, OR, NOT statements)

Note:

If we want to look for patterns i.e. we want to look for all the names that start with ‘Ch’, we can

WHERE

Field1 = ‘Ch%’

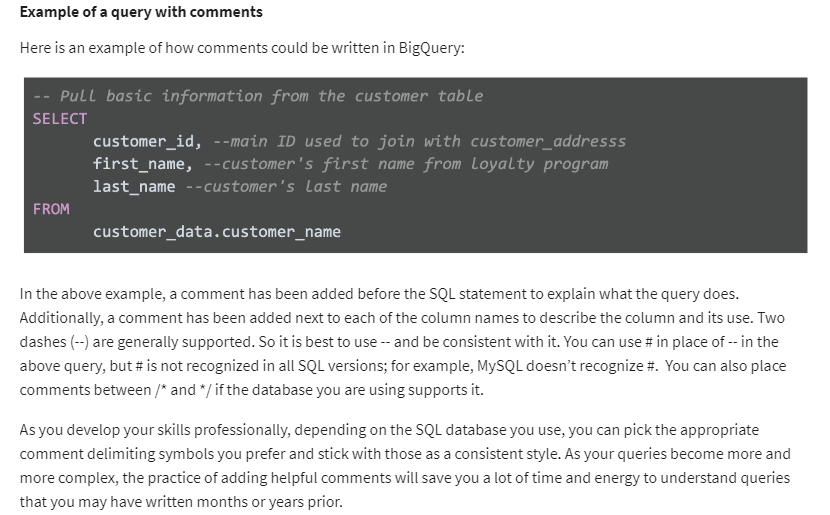
Note that sometimes the percent sign is replaced with an asterisks sign – called a wildcard

Note:

To write a comment

/\* comment here \*/

-- double dashes also allows you to insert a comment



Aliases

You can rename things using aliases using the AS clause. It doesn’t change the actual name of the column or table in the database, but you can use it in your queries eg.

Field1 AS last\_name

Table AS customers

W3Schools SQL Tutorial

[SQL Tutorial (w3schools.com)](https://www.w3schools.com/sql/default.asp)

SQL Cheat Sheet

[SQL Cheat Sheet. Starter guide for standard SQL syntax… | by Jason Lee | Towards Data Science](https://towardsdatascience.com/sql-cheat-sheet-776f8e3189fa)

**Data Visualization**

Steps to plan a data visualization:

1. Explore the data for patterns

* Ask the manager or data owner for access to data
* Gather information relevant to the problem
* Notice any patterns
  + Eg. Do people from particular regions visit the website more frequently, do people in certain age ranges spend more?

1. Plan your visuals

* Refine the data and present the results of your analysis
  + Eg. For a sales oriented audience
    - Show sales numbers over time
    - Connect sales to location
    - Show the relationship between sales and website uses
    - Show which customers fuel growth

1. Create your visuals

* Use different graphs or charts for different information
  + Eg. Line charts can track sales over time
  + Eg. Maps can connect sales to locations
  + Eg. Donut charts (pie graphs) can show customer segments
  + Eg. Bar charts can compare total visitors to visitors that make a purchase

Many options for visualization:

* Spreadsheets – simple, quick, easy to use
* Tableau – visualization software, much more powerful
  + [Resources | Tableau Public](https://public.tableau.com/en-us/s/resources)
* Rstudio – programming language