

# PowerQuery - a practical introduction

Brendan Clarke, NHS Education for Scotland, [brendan.clarke2@nhs.scot](mailto:brendan.clarke2@nhs.scot)

17/07/2024

# Welcome

- 🌂 this session is a beginner-level introduction to PowerQuery
- we'll get going properly at 14.35
- you'll either Excel desktop (>2007) or Power BI desktop
- you'll also need to have basic competency in Excel or Power BI
- if you can't access the chat, you might need to join our Teams channel:  
[tinyurl.com/kindnetwork](https://tinyurl.com/kindnetwork)
- you can find session materials at [tinyurl.com/kindtrp](https://tinyurl.com/kindtrp)

# The KIND network

- a social learning space for staff working with knowledge, information, and data across health, social care, and housing in Scotland
- we offer social support, free training, mentoring, community events, ...
- Teams channel / mailing list

# Excel / Power BI training sessions

Session	Date	Area	Level
Lambda formulas in Excel	13:00-13:30 Mon 15th July 2024	Excel	🔪🔪 : intermediate-level
Lookups in Excel	13:00-14:30 Thu 1st August 2024	Excel	🔪🔪 : intermediate-level
Relative, absolute, mixed, structured, and R1C1 references in Excel	15:00-16:00 Thu 8th August 2024	Excel	🔪🔪 : intermediate-level

- this session is based on our **Power BI for Beginners** course
- but PowerQuery isn't specific to Power BI - it's brilliant for Excel too

# Session outline

- what is PowerQuery, and why might you use it?
- getting data into PowerQuery
- basic data wrangling
- joins (merges)
- pivots

# What is PowerQuery?

- Microsoft call it a “data transformation and data preparation engine”
- it’s been around since 2010, initially in Excel, and then in both Excel and Power BI
- uses code (M Formula language) but has GUI tools for most tasks

# How I think of PowerQuery

- the right way to load any real data into Excel

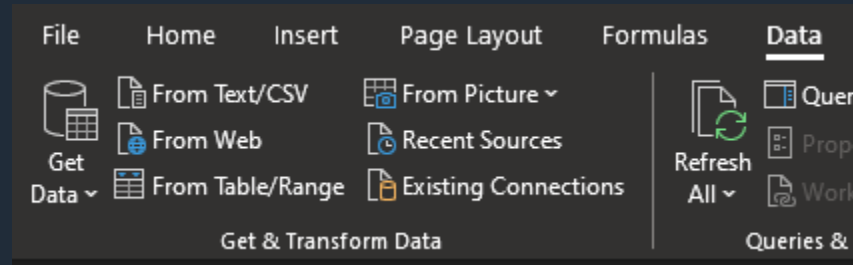


# Why might you use it?

- powerful
- easy, to start with at least...
- allows you to standardise data loading from different sources
- allows you to repeat data loading

# Getting data into PowerQuery (Excel)

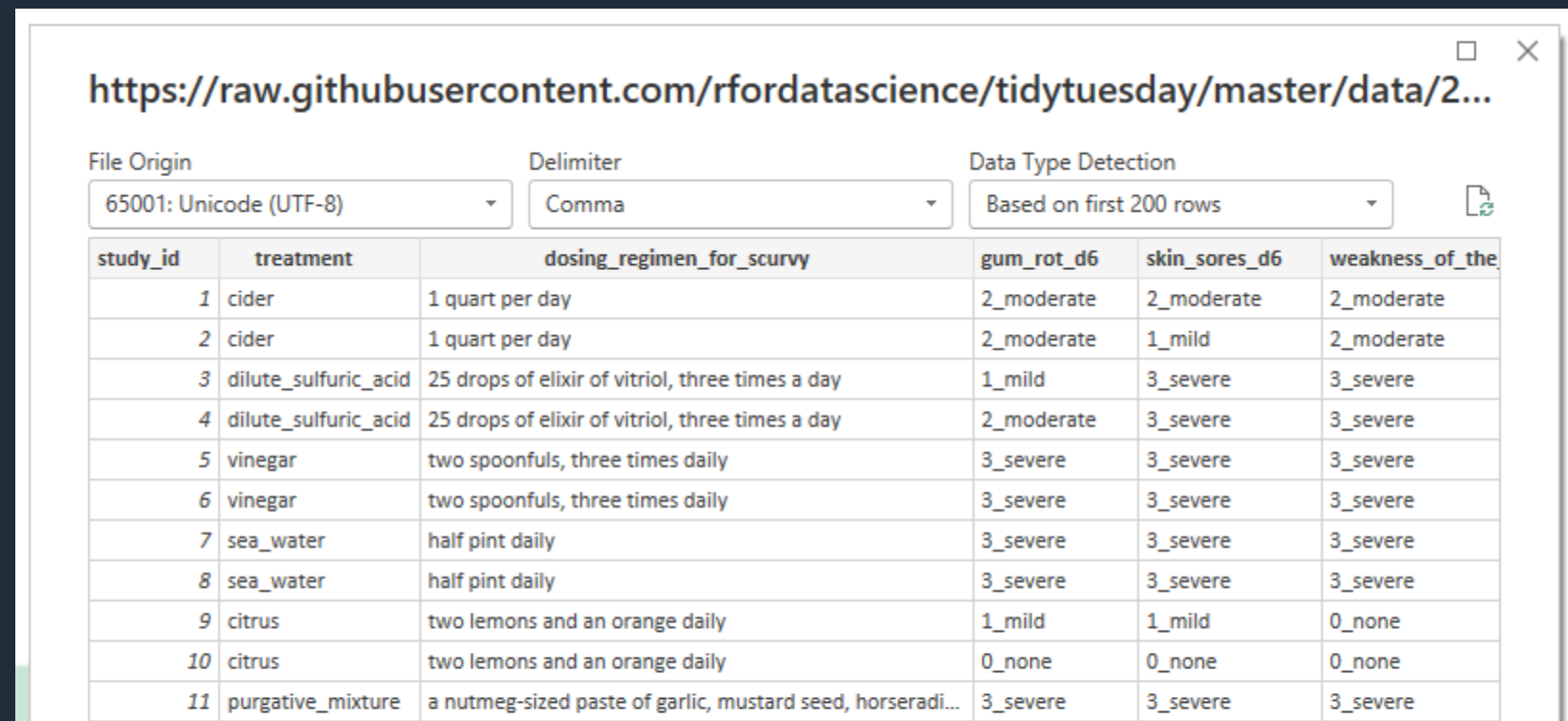
1. Open Excel and create a new, blank, workbook
2. **Data > Get & Transform Data**



3. Pick **From Web** and supply the following URL -  
[https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020-07-25/scurvy.csv](https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-07-25/scurvy.csv)
4. From the preview, select **Transform data** to open PowerQuery

# Getting data into PowerQuery (Power BI)

1. Open Power BI and start a new report
2. Select **Get data > From Web** button in the ribbon
3. Supply the following URL -  
[https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020-07-25/scurvy.csv](https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-07-25/scurvy.csv)
4. From the preview, select **Transform data** to open PowerQuery



The screenshot shows the 'Web Content' preview window in Power BI. The title bar displays the URL: <https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020-07-25/scurvy.csv>. Below the title bar, there are three dropdown menus: 'File Origin' set to '65001: Unicode (UTF-8)', 'Delimiter' set to 'Comma', and 'Data Type Detection' set to 'Based on first 200 rows'. The main area displays a table with the following data:

study_id	treatment	dosing_regimen_for_scurvy	gum_rot_d6	skin_soers_d6	weakness_of_the
1	cider	1 quart per day	2_moderate	2_moderate	2_moderate
2	cider	1 quart per day	2_moderate	1_mild	2_moderate
3	dilute_sulfuric_acid	25 drops of elixir of vitriol, three times a day	1_mild	3_severe	3_severe
4	dilute_sulfuric_acid	25 drops of elixir of vitriol, three times a day	2_moderate	3_severe	3_severe
5	vinegar	two spoonfuls, three times daily	3_severe	3_severe	3_severe
6	vinegar	two spoonfuls, three times daily	3_severe	3_severe	3_severe
7	sea_water	half pint daily	3_severe	3_severe	3_severe
8	sea_water	half pint daily	3_severe	3_severe	3_severe
9	citrus	two lemons and an orange daily	1_mild	1_mild	0_none
10	citrus	two lemons and an orange daily	0_none	0_none	0_none
11	purgative_mixture	a nutmeg-sized paste of garlic, mustard seed, horseradi...	3_severe	3_severe	3_severe

# Preview your Data

<https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2023/2023-07-25/scurvy.csv>

File Origin: 65001: Unicode (UTF-8) | Delimiter: Comma | Data Type Detection: Based on first 200 rows

study_id	treatment	dosing_regimen_for_scurvy	gum_rot_d6	skin_soers_d6	weakness_of_the_knees_d6	lassitude_d6	fit_for_duty_d6
1	cider	1 quart per day	2_moderate	2_moderate	2_moderate	2_moderate	0_no
2	cider	1 quart per day	2_moderate	1_mild	2_moderate	3_severe	0_no
3	dilute_sulfuric_acid	25 drops of elixir of vitriol, three times a day	1_mild	3_severe	3_severe	3_severe	0_no
4	dilute_sulfuric_acid	25 drops of elixir of vitriol, three times a day	2_moderate	3_severe	3_severe	3_severe	0_no
5	vinegar	two spoonfuls, three times daily	3_severe	3_severe	3_severe	3_severe	0_no
6	vinegar	two spoonfuls, three times daily	3_severe	3_severe	3_severe	3_severe	0_no
7	sea_water	half pint daily	3_severe	3_severe	3_severe	3_severe	0_no
8	sea_water	half pint daily	3_severe	3_severe	3_severe	3_severe	0_no
9	citrus	two lemons and an orange daily	1_mild	1_mild	0_none	1_mild	0_no
10	citrus	two lemons and an orange daily	0_none	0_none	0_none	0_none	1_yes
11	purgative_mixture	a nutmeg-sized paste of garlic, mustard seed, horseradi...	3_severe	3_severe	3_severe	3_severe	0_no
12	purgative_mixture	a nutmeg-sized paste of garlic, mustard seed, horseradi...	3_severe	3_severe	3_severe	3_severe	0_no

Buttons: Load | Transform Data | Cancel

(make sure to click **Transform data**)

# First view of PowerQuery

The screenshot displays the Microsoft Power Query interface. At the top is a ribbon menu with various options such as 'Close & Load', 'Refresh', 'Advanced Editor', 'Choose Columns', 'Remove Columns', 'Keep Rows', 'Remove Rows', 'Sort', 'Split Column', 'Group By', 'Data Type: Whole Number', 'Merge Queries', 'Append Queries', 'Combine Files', 'Manage Parameters', 'Data source settings', 'New Source', 'Recent Sources', and 'Enter Data'. Below the ribbon is a data preview table with columns: 'study\_id', 'treatment', 'dosing\_regimen\_for\_scurvy', 'gum\_rot\_d6', 'skin\_sores\_d6', 'weakness\_of\_the\_knees\_d6', and 'lassitude\_d6'. The table contains 12 rows of data. On the right side, the 'Query Settings' pane is open, showing the 'Name' as 'scurvy' and the 'APPLIED STEPS' list, which includes 'Source', 'Promoted Headers', and 'Changed Type'.

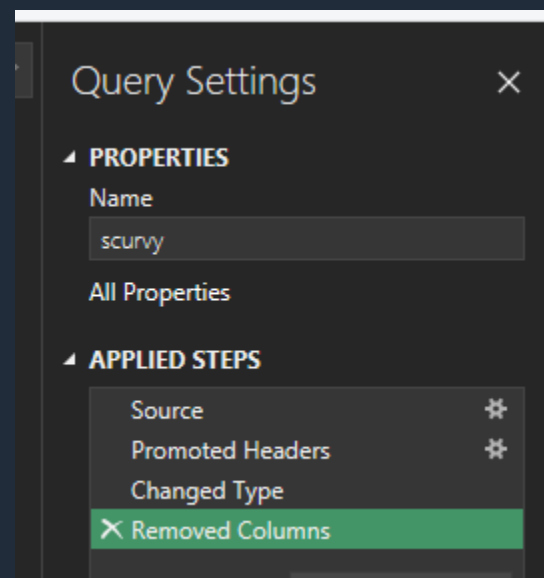
	study_id	treatment	dosing_regimen_for_scurvy	gum_rot_d6	skin_sores_d6	weakness_of_the_knees_d6	lassitude_d6
1	1	cider	1 quart per day	2_moderate	2_moderate	2_moderate	2_moderate
2	2	cider	1 quart per day	2_moderate	1_mild	2_moderate	3_severe
3	3	dilute_sulfuric_acid	25 drops of elixir of vitriol, three times a day	1_mild	3_severe	3_severe	3_severe
4	4	dilute_sulfuric_acid	25 drops of elixir of vitriol, three times a day	2_moderate	3_severe	3_severe	3_severe
5	5	vinegar	two spoonfuls, three times daily	3_severe	3_severe	3_severe	3_severe
6	6	vinegar	two spoonfuls, three times daily	3_severe	3_severe	3_severe	3_severe
7	7	sea_water	half pint daily	3_severe	3_severe	3_severe	3_severe
8	8	sea_water	half pint daily	3_severe	3_severe	3_severe	3_severe
9	9	citrus	two lemons and an orange daily	1_mild	1_mild	0_none	1_mild
10	10	citrus	two lemons and an orange daily	0_none	0_none	0_none	0_none
11	11	purgative_mixture	a nutmeg-sized paste of garlic, mustard seed, horseradish, balsam of P...	3_severe	3_severe	3_severe	3_severe
12	12	purgative_mixture	a nutmeg-sized paste of garlic, mustard seed, horseradish, balsam of P...	3_severe	3_severe	3_severe	3_severe

## First view of PowerQuery

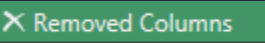
- Queries (= datasets) on the left
- ribbon menu at the top
- Query settings (= the changes you'll make to your data) on the right
- Excel-ish data preview in the middle

# Basic data wrangling

- right click the `study_id` column and select **Remove**
  - the column will disappear
  - but you'll also gain a step in your Applied Steps field



# Basic data wrangling

- you can edit this (and any) step
  - if you're code-confident, you can edit the formula to read = `Table.RemoveColumns("#Changed Type", {"study_id", "treatment"})` to also remove the treatment column
  - and you can remove the step using the X next to the step name 
- you should also try renaming and reordering your columns in the same way
- once you've had a play, please remove the steps until you're left with the data you started with (so delete back to "Changed Type")

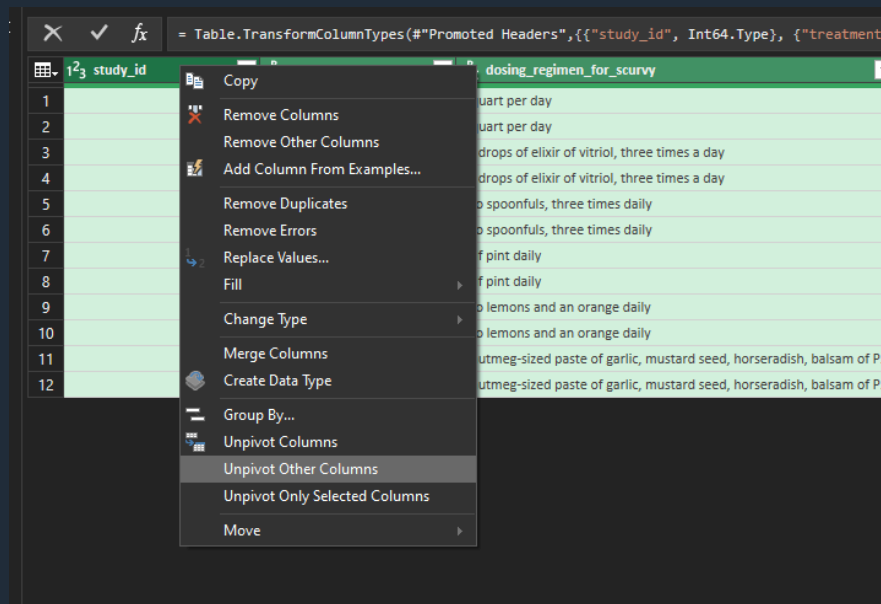


# Replacing values

- we can also work with individual values
- our values are a number, then an underscore, then a word
- we can **Split column** at the underscore
- we could also replace the odd underscores in the treatment column

# Reshaping data

- we could go column by column here splitting, but an easier approach would be to re-shape the data
- select the first three columns, and select **Unpivot Other Columns**

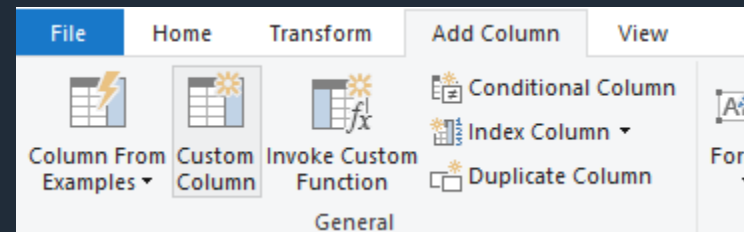


- then split, change type, and generally tidy up

# Add columns

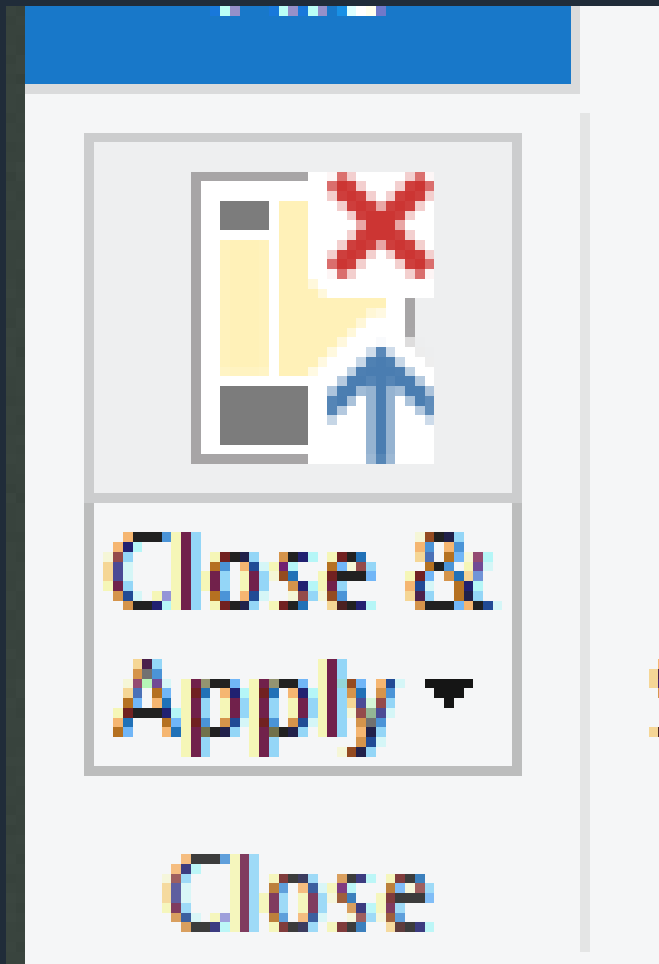
Our treatment and dosing columns are annoying. Let's make a new, joined, column for them:

- select **Custom Column**



- you can click the gearwheel next to the step if you'd like to go back and tweak

# Close and load / Close and Apply sends your data back to Excel/Power BI



# Getting more fancy: a bigger dataset

- please load the Health and Care Experience survey results from 2023-4

The screenshot shows a web application interface. On the left is a 'Navigator' panel with a search bar, a 'Select multiple items' checkbox (unchecked), and a 'Display Options' dropdown. Below these are three items: 'Background', 'Information Questions', and 'Positive, Neutral or Negative'. The 'Information Questions' item is highlighted in green. On the right is a table titled 'Information Questions' with the following data:

Geography Type	Area	Area Name	Survey Section
Scotland	Scotland	Scotland	Your General Practice
Scotland	Scotland	Scotland	Your General Practice
Scotland	Scotland	Scotland	Your General Practice
Scotland	Scotland	Scotland	Your General Practice
Scotland	Scotland	Scotland	Your General Practice
Scotland	Scotland	Scotland	Your General Practice
Scotland	Scotland	Scotland	Your General Practice
Scotland	Scotland	Scotland	Your General Practice
Scotland	Scotland	Scotland	Your General Practice
Scotland	Scotland	Scotland	Your General Practice

- or alternative local version
- please load both the proper datasets (and ignore the background)

This screenshot shows the 'Navigator' panel with the 'Select multiple items' checkbox checked. The 'Background' item has an unchecked checkbox, 'Information Questions' has a checked checkbox, and 'Positive, Neutral or Negative' has a checked checkbox. The 'Positive, Neutral or Negative' item is highlighted in green.

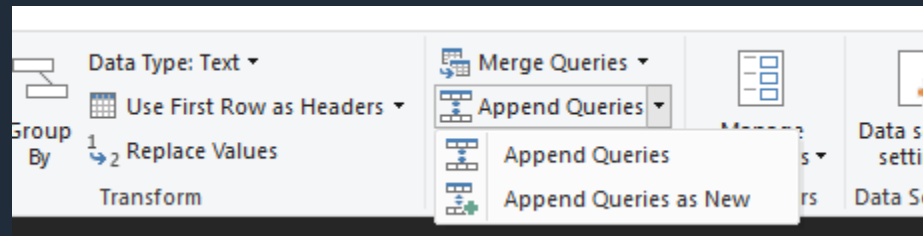
# A challenge

Can you find the number of responses to each question for your local NHS board? You should be able to re-work this data to give something like:

	ABC Survey Section	ABC Question Number	ABC Question Text	ABC 123 Number of Responses
1	About You	q38	In general, how well do you feel that you are able to look after your o...	6650
2	About You	q39	Do you have any of the following, which have lasted, or are expected t...	6493
3	About You	q40	Are your day-to-day activities limited because of a health problem or d...	6622
4	About You	q41	Do you suffer from chronic or persistent pain, that is pain that carries ...	6544
5	About You	q42	Which of the following best describes your sexual orientation?	6456
6	About You	q43	What best describes your ethnic group? Please tick one box only	6638
7	About You	q44	What religion, religious denomination or body do you belong to?	6571
8	Care, Support and Help with ...	q27	In the last 12 months, have you had any help or support with everyday...	6402
9	Care, Support and Help with ...	q28	Who funds your help or support with everyday living? Please tick all th...	957
10	Care, Support and Help with ...	q29	Which of the following applies to you and how your social care is arran...	740
11	Care, Support and Help with ...	q22	If you are not receiving all the help and support you need for your daily li...	955

# Some hints

- filtering works like Excel, but is destructive
- unpivoting, doing something to all the rows, then re-pivoting is a great way to work across lots of columns
- you can join structurally similar datasets using **Append Queries as New**



- you can deduplicate entire datasets by selecting all the columns and using **Remove Rows > Remove Duplicates**



# Feedback and resources

- please can I ask for some feedback - takes less than a minute, completely anonymous, helps people like you find the right training for them

Session	Date	Area	Level
Lambda formulas in Excel	13:00-13:30 Mon 15th July 2024	Excel	🌶️🌶️ : intermediate-level
Lookups in Excel	13:00-14:30 Thu 1st August 2024	Excel	🌶️🌶️ : intermediate-level
Relative, absolute, mixed, structured,	15:00-16:00 Thu 8th August 2024	Excel	🌶️🌶️ : intermediate-level

Session

Date

Area

Level

---

and R1C1 references  
in Excel