# Data Acquisition and cleaning exercise

**Exercise 1:** Open the file and clean the data.

1. Download the file *WFH-Survey-Responses-NSW.xlsx* (Edited version from original 2020 survey found at https://data.nsw.gov.au/data/dataset/nsw-remote- working-survey).
2. Open the file with Microsoft Excel Live.
   a. Open your University email account.
   b. Click on the blue cloud icon from the left-hand side menu. This will open OneDrive on a new tab.
   c. On the top menu, click on *Upload.* A pop-up window will open to search for the file you want to upload to OneDrive. Navigate to the folder where you saved the file you downloaded from Canvas and select it.
   d. The file will now be uploaded to your OneDrive.
   e. Wait until the file is uploaded, and click on it once it appears on the lists of files. This will open the file on Excel Live.
3. Look at the responses and see if you need to clean the data.

*Some observations:*

- Some responses have *NA* or an empty cell when the person didn't respond to that specific question. When cleaning the data, you can decide to keep all those responses, fill the empty cells with *NA*, and count that as a possible answer. Otherwise, you can remove the whole row of responses. It all depends on what your goal is and if the number of affected rows would have a great impact on the result.
- Some questions have responses in different formats. You need to choose a preferred format and update all the responses to that format.
- Multi-valued data (comma separated values).

**Exercise 2:** Summarising Ordinal Data.

1. Create a histogram for the question *What year were you born?*
   a. Highlight the cells of the corresponding question.
   b. On the top menu, go to *Insert*. On the *Charts* section, click on *Other Chart.*
   Scroll down to the statistical charts*,* and then select the *Histogram* option. The histogram will automatically appear with default options, and a new *Chart* menu will appear at the top.
   c. From the *Chart* menu, click on *Format.* The chart format options will appear on the right-hand side. Click on *Series*, and you will see the histogram bin options where you can customise the bins to your liking. From the *Bins* drop-

down options, try the *bin width* or the *number of bins* and change the settings to see what happens.

d. Your histogram should look something like this:
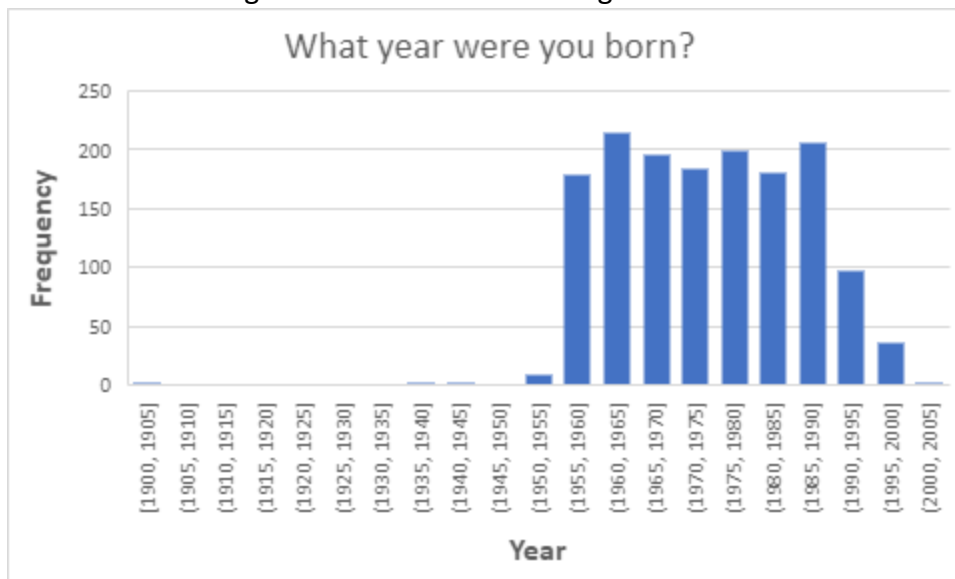


2. Discuss:
   a. Which year were most people born? (*Tip: Select a bin width 1*)
   b. Which decade were most people born?