**Kylie Reynolds**

**FOAR705 - Learning Journal**

**Semester 2, 2019**

Learning journal - General Info

12/08/2019 - 11:31 am

Instructions for this week:

Before class next week, in your learning journal, finish reading to “Formatting Problems”

document the two exercises we’ve (hopefully) done today in your GitHub repository

**Problem**: not sure how to do this or what I am meant to be documenting??

try and find out through looking through slack comments/questions or relistening to echo recording, if no answer found, ask on Slack

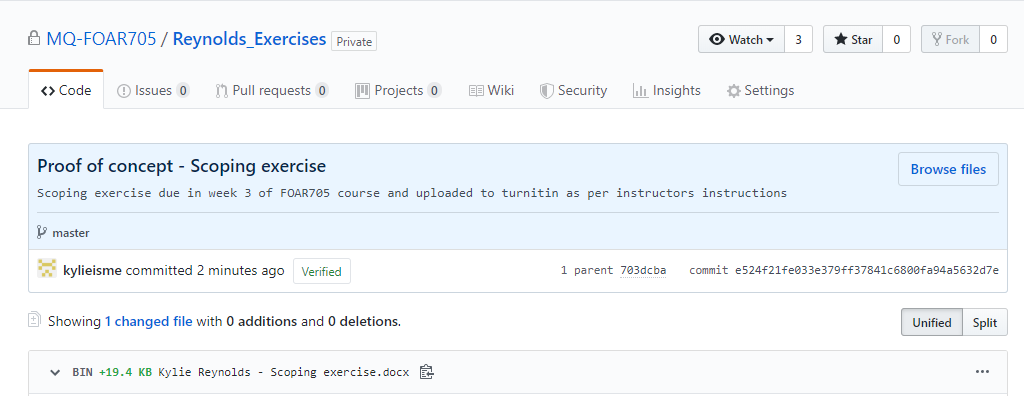
**Solution**: Notes from Brian on Slack for journal entry (09/08/2019, 7:57 pm):

All technical work outside of class should be recorded in a laboratory notebook (document kept with the code or on Cloudstor) which documents the intention of the action, the specifics of the action taken, and the results, along with any marginal notes for improvement or updating your mental model of what should have had happened. This documentation includes: an answer to the specific objective: "What do you intend to be the result of the action you are about to take", the action to be taken containing timestamp, and commands or actions performed, and the result, documenting what happened, success or failure in relation to the objective, and error states. Documentation of errors (each in its own entry) and their remediation are strongly encouraged.

GitHub Committ

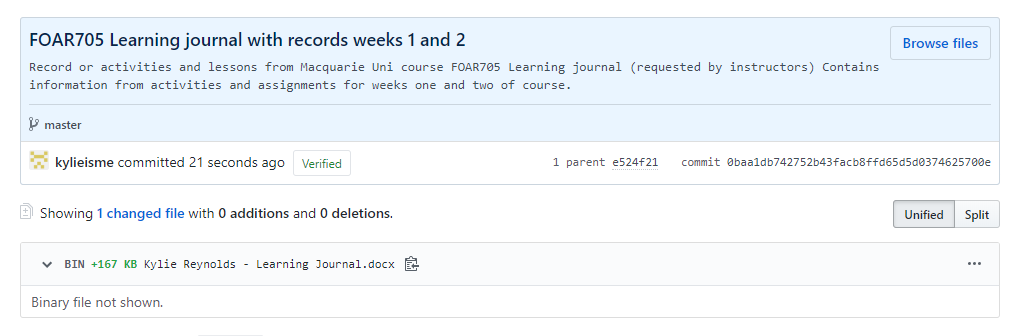
16 August 2019

* Logged in to GitHub - MQ-FOAR705/Reynolds\_Exercises
* Clicked on Upload files
* Uploaded the document from my laptop file
* In Heading text bar for commit comments, I entered the file name which was “Proof of concept - scoping exercise”
* I added a comment to describe the document (see below)
* successfully completed with no errors (as far as I know). Kylie Reynolds - Scoping exercise.docx (commit e524f21fe033e379ff37841c6800fa94a5632d7e) - see below:



GitHub commit:

* Learning journal see below:



# Data Carpentry

**14/08/2019 8:30 pm**

I accessed the Week 2 assignments document on Cloudstor to find out what tasks are assigned for this week. There are two tasks assigned which are:

* Data Carpentry (DC)

**Data Organization in Spreadsheets for Social Scientists**

**Objective**: to read and complete exercises in DC Data Organization in Spreadsheets for Social Scientists (<https://datacarpentry.org/spreadsheets-socialsci/>)

1. Introduction

(<https://datacarpentry.org/spreadsheets-socialsci/00-intro/index.html>).

Three spreadsheets were required for this and subsequent lessons in this section of the course. The spreadsheets were downloaded in the first week under instruction from presenters in the of the Macquarie University (MQ) FOAR705 class. Therefore, I did not need to download the spreadsheets for the lesson today. The spreadsheets are downloadable from: <https://datacarpentry.org/spreadsheets-socialsci/setup.html>

The three spreadsheets that were downloaded are:

* [SAFI\_clean.csv](https://ndownloader.figshare.com/files/11492171)
* [SAFI\_messy.xlsx](https://ndownloader.figshare.com/files/11502824)
* [SAFI\_dates.xlsx](https://ndownloader.figshare.com/files/11502827)
* And they were saved in the “Data Carpentries for Social Scientists/Spreadsheets” folder on my laptop desktop.
* This task was straight forward and the instructions in this lesson were easy to follow and to execute.
* I did not have any problems or come across any errors.

**Reflection**: Maybe it would have been best practice to download the files again today as they may have been updated since last week.

**Solution**: I re-downloaded the spreadsheets and replaced the old ones in the same folder.

The introduction outlined what will be covered in this and subsequent sections of the Data Organization in Spreadsheets for Social Scientists lesson. The main objective of this lesson is to teach us best practice in the organization of data and data entry practices, and the critical importance of the data we collect for our research projects.

* To complete tasks for this lesson, a spreadsheet software is needed. The program that I will be using is excel.
* This section of the lesson lists some of the problems that can arise with spreadsheets. It spells out the usefulness of spreadsheets for data entry purposes, table and figure creation, statistics etc. However, it also warns of some of the dangers associated with these types of processes which could cause problems with research data analysis further on, some of these are:
* White space
* Highlighting of cells
* Merging of cells
* Applying calculations

**Reflection**: I often do all of the things listed above, and now that I understand that this can cause big problems, I will be more aware and mindful of when and how I use these features of the spreadsheet software.

**Solution**: Learning more about good data processing and organisational skills by continuing through the next lessons will help.  And discontinuing bad practices and habits which may compromise my research data.

**Exercises**:

* How many people have used spreadsheets in their research?
  + I assume most researchers have used spreadsheets in their research to organise and analyse data. I have used spreadsheets to record data for research projects in uni and at work.
* How many people have accidentally done something that made them frustrated or sad?
  + I have accidentally deleted data from an original data source, and I have also organised my data in sloppy ways (as I have learnt today) which has caused problems for me and has most likely caused problems for my co-workers in the analysis stages.

1. Formatting data tables in Spreadsheets

**15/08/2019 8:30 pm - 9:45 pm**

**Objective:** To successfully complete the Data Carpentry for social scientists: Formatting data tables in Spreadsheets module <https://datacarpentry.org/spreadsheets-socialsci/01-format-data/index.html>

This module is teaching us how to avoid making mistakes when working with data and how to avoid creating challenges when formatting of spreadsheets for data collection, storage and analysis.

Common mistakes are:

* adding contextual elements to the data spreadsheet - e.g. margin notes, the layout/design of the spreadsheet, and using fields to convey meaning to contextualise data (e.g. notes, colours etc., I think)
  + may confuse the computer and mess with the validity of calculations and analysis of data

Important:

* good formatting of spreadsheets from the very beginning
* Data organization is the foundation of your research project
* Data entry/analysis tip - automate conversion of files

To keep track of analyses:

* Keep the original dataset and create a new file for any changes/cleaning up of data/analysis
* Track data clean-up/analysis/ step-by-step (experiment note stylz) in a new file or tab (to keep it together)

Structuring data - cardinal rules:

The cardinal rules of using spreadsheet programs for data:

1. Put all your variables in columns
2. Put each observation in its own row.
3. Don’t combine multiple pieces of information in one cell.
4. Leave the raw data raw - don’t change it!
5. Export the cleaned data to a text-based format like CSV (comma-separated values) format.
   * **QUESTION**: how do you do that??

Exercises:

Take a messy version of the SAFI data and describe how we would clean it up.

1. Download the [messy data](https://ndownloader.figshare.com/files/11502824). -
   * Messy data downloaded previously (see notes and links for journal entry 14/08/2019)
   * original files saved on laptop desktop Data Carpentry for Social Scientists, Spreadsheets file.
2. Open the data in a spreadsheet program.
   * I went onto the laptop desktop, double clicked on Data Carpentry for Social Scientists folder, and then Spreadsheets folder, and double clicked on the messy excel file
   * The file auto opened in Excel.
   * I clicked on enable editing button at the top of the page to access the file
   * as this is my raw data file, I then clicked on “File” and then “Save as” and created a copy which I named - SAFI\_messy\_ cleaned\_15082019 in the same folder as the original
   * I moved the raw data SAFI\_messy file into a newly created folder in the Spreadsheets folder, which I named SAFI\_RAW\_data for safe keeping
3. Notice that there are two tabs. Two researchers conducted the interviews, one in Mozambique and the other in Tanzania. They both structured their data tables in a different way. Now, you’re the person in charge of this project and you want to be able to start analyzing the data.
   * I have surveyed both tabs to look at the data and analyse what I am working with
   * My first thoughts are that it would be impossible to analyse the data in this format and that I would need to merge the data and organise it in a way that which makes sense and according to the good data organization principles learnt in this module and in the introduction module
4. Identify what is wrong with this spreadsheet. Discuss the steps you would need to take to clean up the two tabs, and to put them all together in one spreadsheet.

|  |  |
| --- | --- |
| What is wrong | Steps to be taken |
| Data separate - needs to be merged | Merging of data:   * Create new tab * create columns for each variable * each observation must have own line * Carefully and correctly enter the data into the new sheet * Double-check data entry against raw data source   Possible problem: not sure how to deal with the Key\_IDs as they are, they should be different for each entry and there will be duplicates if using the original data from both locations)  Possible solution: (need to find out if it is ok to do this??) renumber the key\_ID and make a note in a notes section which will be created in a separate tab to log what has been done, step-by-step |
| Contextual information and formatting that may cause analysis problems - e.g. White space, colour coding, notes in random places | * in new spreadsheet cells only to be used for recording data and column names * no merging of cells * add a column for “notes” - any notes from observation can be put there * Any random notes or colour coding etc. can be logged in a separate notes tab or file * Merged cells and white space for design purposes are a no-no |
| The title of the study and “2017 Data Collection” info | This information can go in the Notes tab. |
| Blank A column | Organise and format according to Tidy Data principles:  1. Each variable forms a column.  2. Each observation forms a row.  3. Each type of observational unit forms a table (Wickham, 2014, p. 4).  The first column should start with Key\_ID to identify the observation line and all other columns should continue from there with no missed columns in between |

Metadata exercise:

* What is not immediately obvious to me about this data?

I opened the previously downloaded copy of SAFI\_clean spreadsheet saved on laptop desktop Data Carpentries for Social Scientists, Spreadsheets folder and noticed that some of the variable column titles are not self-explanatory. There are also no notes connected to the document and no references to any other files which could explain the meanings of the variable titles,values, or the questions that were asked which could give me some context information which I would need when thinking about during the analysis of the data.

* What questions would I need to know the answers to in order to analyze and interpret this data?
* I would need to know exactly what the column headers stand for
* I would need to know what was observed, how it was observed, the type of data that was required for each response, how data was coded, description of each variable, value meanings, etc.

1. Formatting problems

**16/08/2019 9:30 am - 9:55am**

**Objective:** To successfully complete the Data Carpentry for social scientists: Formatting problems module <https://datacarpentry.org/spreadsheets-socialsci/02-common-mistakes/index.html>

Common spreadsheet formatting problems/solutions:

* Using multiple tables - Don’t do it - it confuses the computer, which may not be able to read associations between the tables
* Using multiple tabs - Data split between tabs - don’t do it - computer cannot make connections
* Not filling in zeros - empty spaces means no data collected (null) - zero are entered if questions has been answered and the value is zero
* Using problematic null values - Blanks (most applications) and NA (for R) are best to use - check for different software
* Using formatting to convey information - do not highlight or leave blank rows in data - don’t use design features in data tables, e.g. merging cells etc.
* Using formatting to make the data sheet look pretty
* Placing comments or units in cells - do not write notes in cells with other units of data - create a new notes field
* Entering more than one piece of information in a cell - each cell should only have on unit of data - create new columns instead
* Using problematic field names - Yes: wall\_type OR WallType - No: wall type - no spaces and clear naming instead of abbreviations that may not make sense down the track
* Using special characters in data - Word, formatting and fancy non-standard characters can mess up data - be careful when copying and pasting data and do not use.

**Reflection**: The common formatting problems that are listed and demonstrated on the SAFI\_messy spreadsheet are many. When I picked up problems for the last module exercises, I seemed to have missed quite a few of the mistakes that were.

**Solution**: create a checklist of problems to be mindful of when formatting spreadsheets for research data and to check when I have been supplied data from another person.

1. Dates as data

21/08/2019 9:30pm

**Objective**: To successfully complete the Data Carpentry for social scientists: Dates as data module <https://datacarpentry.org/spreadsheets-socialsci/03-dates-as-data/index.html>

General notes:

* Single column for dates is not best practice
  + may cause problems for computer
  + allow ambiguity to creep into your data
* functions differ between spreadsheet programs - may not be compatible if exporting

Dates as integers (whole numbers not fractions of):

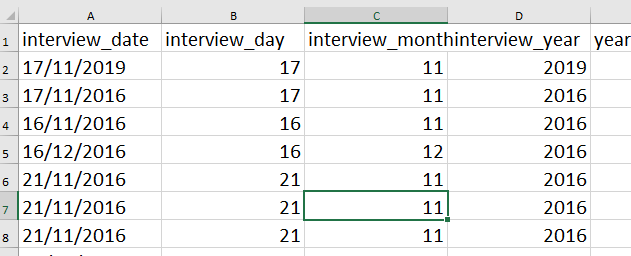
* Excel counts date from December 31, 1899 and stores numbers as serial numbers (e.g. 41834 same as 07/14/14)
* “ [dates must be checked for accuracy when exporting data from Excel](http://datapub.cdlib.org/2014/04/10/abandon-all-hope-ye-who-enter-dates-in-excel/)” (this link in Data carpentry lesson doesn’t work
* Formula for adding data in Excel e.g. =B2(cell with date)+90(days) = new date

Regional date formatting:

* different countries format dates differently (e.g. US month/day backwards) - this is why should not put dates in same column
* put in separate columns - 3 pieces of data (month, day, and year)
  + avoids confusion
  + better for comparison

Exercise:

* download spreadsheet - SAFI\_dates.xlsx - https://ndownloader.figshare.com/files/11502827 (already completed this task in first week of class) file saved in DataCarpentry/Spreadsheets folder
* I opened the spreadsheet and pressed the enable editing button at the top of the page
* I created three new columns (B, C, D) on the right side of the date column (column A where dates are formatted as 17/11/2016)
* I selected columns B, C, and D, and formatted the columns to numbers
* I named column B interview\_day, C interview\_month, D interview\_year
* In the first row of Column B i entered =DAY(A2) and pressed enter, it showed up as 17.00
* I selected Column B C D and reduced the decimal places to zero
* In the first row of Column C I entered =MONTH(A2) and pressed enter and it showed up as 11
* In the first row of Column D I entered =YEAR(A2) and it showed up as 2016
* Default year exercise - in Cell A2 I changed the date to 17/11 and the year changed to 2019 (default year is the current one)



Historical data (beware):

* pre 31 Dec 1899 Excel will leave as is
* Be very careful when mixing before and after dates

1. Quality assurance

22/08/2019 8:37pm

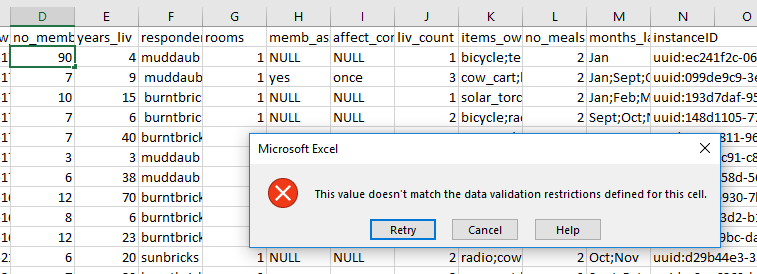
**Objective:** To successfully complete the Data Carpentry for social scientists: Quality assurance module <https://datacarpentry.org/spreadsheets-socialsci/04-quality-assurance/index.html>

Validate data on input

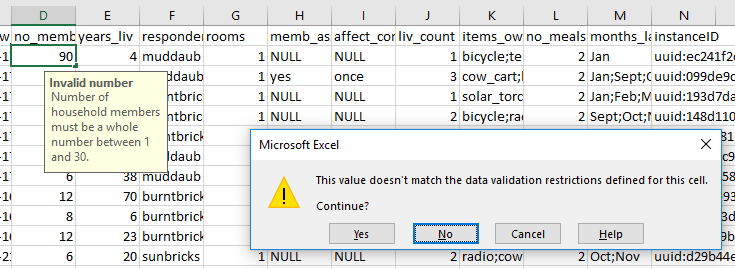
* one type of data per column
* it is possible to add data restriction for values in columns and cells
* In Excel can apply data validations to cells which can raise errors to alert us and data is not entered
* In Excel, can also add validation criteria to cells with data already entered - data not removed and it is flagged by a triangle in top left of cell
* Excel validation rules available here: <https://support.office.com/en-us/article/Apply-data-validation-to-cells-29FECBCC-D1B9-42C1-9D76-EFF3CE5F7249>

Restricting data to a numeric range

* I opened the clean SAFI file downloaded in previous lesson
* Checked column D as per instructions - numbers in this column should be integers and amount should be limited (in this instance 30 is the limit)
* I selected the Column D (no\_members)
* I located the Data tab and selected Data tools and then Data validation and a pop up appeared
* I pressed the drop-down menu arrow on the right of the 1st drop down box
* I selected Whole number and entered Min. 1 and Max 30 as the range
* to test it out I entered 90 into the cell D2 and a pop up populated (see below)



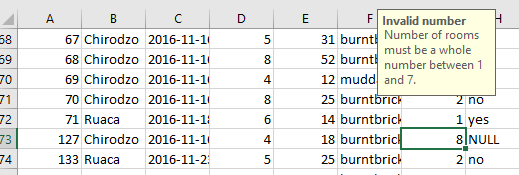
* I pressed cancel to get rid of the error message
* selected Column D again and went to the data validation tool again
* in the pop up I selected Input message tab - types in the message to instruct user why the error had populated, and the rule applied to guide them in how to enter the correct data (1-30 number range)
* I then selected the Error alert tab and changed the Style to Warning.
* I then tried to re-enter the incorrect number in cell D2 to see if it was working (see below)



Exercise:

Apply a new data validation rule to a numeric column

* Using the same clean SAFI file, I selected column G (rooms)
* I located the Data tab and selected Data tools and then Data validation and a pop up appeared
* I pressed the drop down menu arrow on the Settings drop down and I selected Whole number and entered Min. 1 and Max 7 (because there is one entry where the number is 8 and I want to test to see if the triangle appears to let me know that the data is incorrect)
* I then selected the Input message tab entered Invalid number in the title and “Number of rooms must be a whole number between 1 and 7.” in the Input message
* I selected the Error Alert tab and changed the Style setting from Stop to Warning and pressed OK
* to test it out I entered 90 into the cell G2 and a pop up populated - the warning populated, and I pressed cancel and the original data remained
* I checked the cell where the number 8 was (G73) however no little triangle populated to tell me there was a validation error (hmmm..... Y?), however, there is an annoying little pink square (see below) which was sitting over the data at the top of the column which I can’t get rid of, is that trying to tell me there is something wrong or is that just what happens when you add a validation rule to the column.



* To test the annoying box theory, I am going to change the 8 to a 3 to see if the box disappears
* The box did not disappear. Which is annoying but handy as I am unlikely to add the incorrect data with it in my face.
* Aha - Just found the tip at the bottom of Data Carpentry lesson that says if it is an existing spreadsheet with data already in it and it breaks the new validation rule it will not be flagged - that answers my question. Note to self: read the whole lesson so you don’t waste time on unnecessary experiments when the answer is already there. Though I am more likely to remember that now, so all is well in the world.

1. Exporting data

22/08/2019 9:54pm

**Objective:** To successfully complete the Data Carpentry for social scientists: Exporting data module <https://datacarpentry.org/spreadsheets-socialsci/05-exporting-data/index.html>

* Don’t save files in .xls or xlsx
  + may cause problems in the future or may not be able to open in other software or other versions of Excel
  + Some data repositories may not accept this file format
* Save files in universal, open and static format like TSV or CSV
  + “tab-delimited (tab separated values or TSV) or comma-delimited (comma separated values or CSV”
  + can use nearly any software to open and view
  + importation easy for other uses
  + may give you a warning when saving but that’s ok

Exercise:

* Click File and Save As
* in File format section select CSV
* Check save location and click Save
* Success!

Note: check for commas before saving in CSV or enclose data including commas in quotation marks.

# Overleaf

 13/08/2019 10:00 pm

Create an Overleaf account - <http://overleaf.com/edu/macquarie>

I clicked on the link to the Macquarie University link on Overleaf which was posted on the FOAR705 Slack chat (instant messaging and collaboration tool) that @Brian had put up. It is suggested that assignment and learning journal submissions be in LaTex format, and I believe this website will allow me to create this type of document. I believe that LaTex is a typesetting design program but am not sure. ACTION: find out what it does and how to operate or create documents.

I set up my sign in using my Macquarie Uni HDR ID, an email was sent to my email account for verification. Verification was actioned, and this took me to a profile page where I added details of my department (Sociology) and the degree that I am enrolled in (Masters). I noticed that I could link the Overleaf account to my GitHub account which I did, however, I am unsure what this does or how this works. I received an email to notify me that it the GitHub account was successfully linked. ACTION: I will need to find out about this.

I found a button on the webpage which indicated that I could start a project. I clicked on the button where some options came up to set up a blank document, certain other types of documents, or to select a template. First, I clicked on the templates link and had a quick look at what was available. I then decided to click back and to select a blank document. The blank page came up with some details on it. The details looked like code. I clicked on a PDF button on the left side of the page and it showed me a readable version of the text which was coded on the previous link. I went back to the code link and didn’t know what to do so closed the app.

**Reflection**: I need to discover how to use Overleaf.

**Solution**: I looked up “adding text to a document in overleaf” on google and clicked on a video that showed up in the search “LaTeX video tutorial for beginners (video 1) - <https://www.overleaf.com/learn/latex/LaTeX_video_tutorial_for_beginners_(video_1)> to learn how to use the program

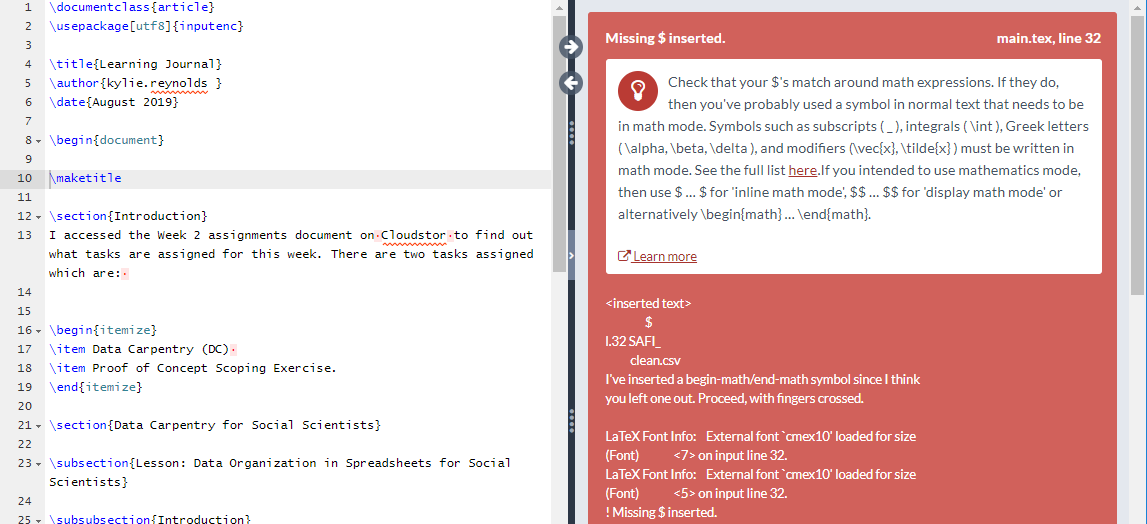
14/08/2019 10:25 pm - Watched video 1 and 2 about basics and adding text to LaTex document - need to trial this process out.

**Reflection**: Note sure if Overleaf is called a program

**Solution**: need to find out about the lingo - ask for help in Slack chat, or at consultation on Friday.

**16/08/2019 12:00 pm**

I attempted to create a learning journal in Overleaf, and things seemed to be going well with me copying and pasting from a text document. I then recompiled the document to see what it would look like and I noticed a little note on top of the viewing panel, which I hovered over, and which said that it was a “Logs and output file”. I opened the note by clicking on it and a whole bunch of information that I could not understand showed up. See below for an example:

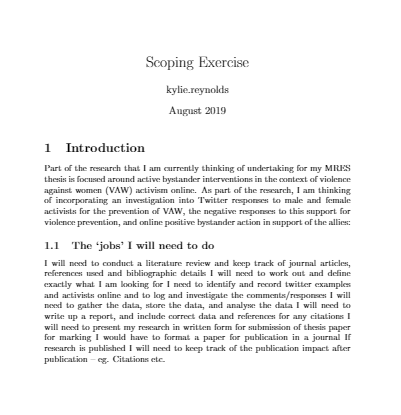


This overwhelmed me so I logged off and thought maybe next task I will feel more confident, I do not know how to deal with this information. I will attempt to do the scoping exercise on Overleaf instead.

18/08/2019 9:33 am

**Objective**: to create an Overleaf document for my Proof of concept: Scoping exercise

1. I logged on to Overleaf
2. Added a new project
3. Named it Scoping Exercise and it automatically took me to the document.
4. I copy and pasted the Introduction paragraph from my word document into the already set up Introduction section
5. I clicked on recompile on the viewing screen and the text seems to have translated properly.
6. I then added a subsection and named it “The jobs I will need to do”
7. I copy and pasted a list of jobs underneath this heading
8. I clicked the recompile button to see on the viewing page
9. **Problem**: The text showed up in a paragraph instead of a list. See image below.
10. **Solution**: I need to figure out how to change this by learning the command codes for creating a bullet list (see below)



18/08/2019 4:19pm

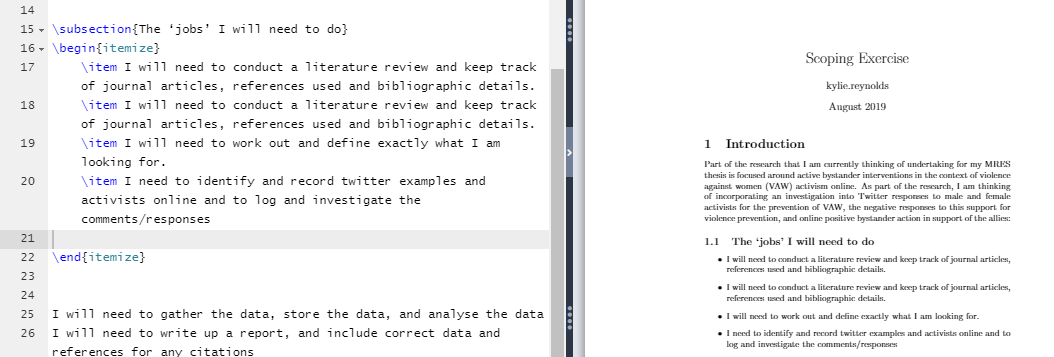
**Objective**:

To learn how to create a bullet list in LaTex

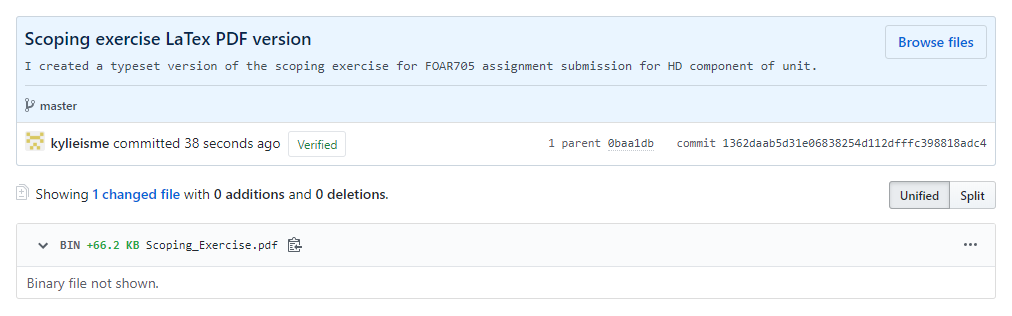
**Steps**:

1. I looked up on Google how to create a bullet list in LaTex
2. One of the links that came up was: <https://latex.org/forum/viewtopic.php?t=12143> which took me to a LaTex Community forum (saved in bookmarks in Uni/FOAR705) question and answer. The answer included the command codes used to create both numbered and bullet lists, which will come in handy.
3. I went to my Scoping exercise project on Overleaf and entered the commands that were presented in the answer referred to above, added a couple of list items and pressed on the recompile button to see if it was successful, and it was. See **Figure 1** below.
4. I will now continue to add the rest of the items to the list, and to create the rest of my subsections, paragraphs, and bullet lists.
5. This was successful and there were no additional problems
6. Once finished I downloaded a PDF version and uploaded to iLearn and Cloudstor.
7. Committ to GitHub, See **Figure 2**.

**Figure 1 -**

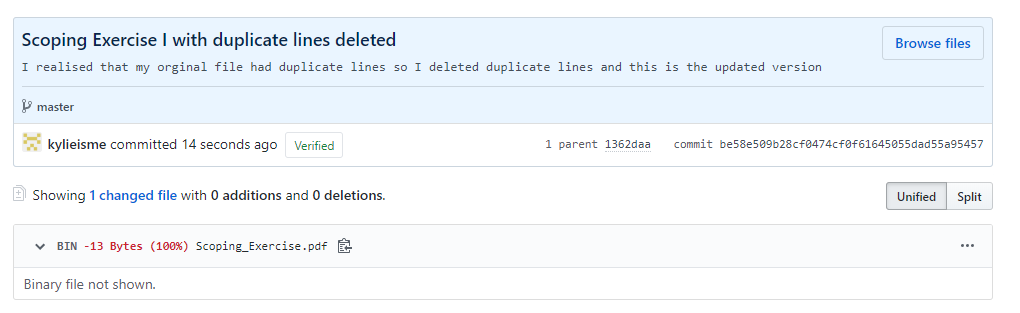


**Figure 2**: GitHub commit for LaTex PDF version of scoping exercise:



21/08/2019 8:09pm

I just had a look at the marks on my scoping exercise I on iLearn and realised that I had doubled up on some bullet points in my LaTex document. I went back in to edit the document and deleted the duplicate lines. I will need to submit on GitHub and will have to remember to check/proofread my work before handing it in.



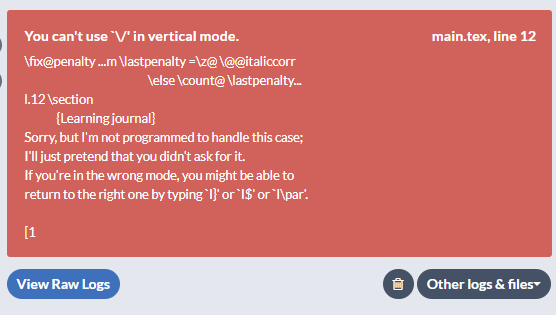
21/08/2019 11:10pm

**Objective**

To typeset my learning journal to prepare for iLearn submission on Friday 23/08/2019

* I had a little trouble with the command codes the other day so I did a google search for LaTex commands and came up with a cheat sheet <https://wch.github.io/latexsheet/latexsheet-a4.pdf>
* I am going to attempt to change the font for my report
* I typed in the code \textsf and then pressed recompile to see if anything changed

**Problem:** an error was populated, see below:



**Solution:**

* I deleted the code and recompiled the document to take it back to how it was originally
* I will need to find out how to change text properly on another day when I have more time (set task in Asana for a reminder)

**Objective:** To reformat the Learning Journal document how I would like it

* First, I need to know how to create a blank line and to get rid of indentations
* On the cheat sheet referred to earlier it says \\ is the code to add a line break, and \noindent to get rid of the indent
* My strategy is to try one thing at a time so that I keep track of what works or what doesn’t work
* I have copied and pasted some unformatted text from word version of learning journal
* I added the \\ in front of the text where I wanted to create the new line, and pressed recompile, but that did not seem to work
* I tried adding the \\ to the end of the line before where I wanted the line added and then pressed enter and then recompile and it worked.
* The line was indented so I entered \noindent at the beginning of the paragraph and it got rid of the indent

Yay!! success.

And now it is time to go to sleep (11:43pm)

22/08/2019 10:35pm

**Reflection**: Turns out last night’s effort was not a success – as self-punishment I have deleted everything and will start again from scratch – this week the journal will have to be submitted in Word format as I don’t have time to learn how to do all that I need to do in LaTex, but, I will not give up!!