Starting out I thought that this module was going to be a nightmare. I felt that it would be too subjective for me, I prefer the technical side of things and this module was one I perceived as being something I just had to struggle through. It is one of my last two modules and that's why I left it to last.

When we went to the initial workshop I had an idea of analysing Facebook feeds to see how the news there differed from traditional media sources. I wanted to go down a route like this as I came at the problem of identifying a research idea from the angle of, 'where can I get data from without causing myself too much trouble!' This is more than likely the wrong approach but having seen the struggles of one or two of my fellow students in getting interviews done I am certainly glad I didn't go down that path. At the workshop I was at a table with John McCarthy and I admitted to him that I approached the research idea from the data gathering point which he said was actually a very reasonable approach. That made me feel a bit better about my plan.

Even though John gave me a bit of guidance about my Facebook idea I eventually settled on analysing Twitter feeds (still looking from a data gathering angle).

My initial idea here was to look at the Twitter feeds of top tennis and golfers and see and differences between the genders. I was going to use a self-presentation framework for this. I was then assigned Mary Galvin as tutor. I met Mary at the qualitative and quantitative workshop. Everyone had to give a brief outline of their research proposal and when I had to give mine I said that I was going to use the self-presentation framework and analyse the data in a quantitative way. Mary then pointed out that it might be interesting to look at the Twitter feeds from a clustering point of view. This would be a qualitative angle and we would let the data lead the way. Qualitative was to me a subjective way of looking at data so I hadn't even considered anything other than quantitative.

The clustering approach would need to be done via R or the Python programming language. As I could program in Python this immediately seemed like a good option.

I went home and wrote some code to get twitter feeds and analyse them. I had read that we were not to start collecting data until we had ethical approval but I was just collecting public data and I wanted to be sure that what I was proposing could be done.

When I had confirmed that I could get the data I realised how much 'cleaning' it would need but that would be something for later.

The next thing was to write up the draft research proposal. I did this and sent it to Mary. She asked why I had picked tennis and golf as the sports to analyse. I had picked them because there was a big difference in their popularity. She pointed out that the issue with attempting to select sports based on that principle is that there's no real evidence to justify the selections. i.e. is there research that justifies the position on the popularity discrepancies in these sports? As she pointed out, it is best to remove as many assumptions as possible. At this point I realised that without the guidance of a tutor I would have been torn to shreds if left to my own devices.

I submitted my research proposal and ethics form and when I received approval I was allowed to proceed.

Along the way I had been reading as many papers as I could for the literature review. This was a massive task as it was hard to know where to stop adding papers to the 'to read' list. It was only when I was faced with the draft literature review deadline that I started to create heading for a rough outline of the review. This was a big help as it focused me on the areas that I needed.

When it came to the references I had planned to use RefWorks due to its ease of keeping references together and putting them in the correct format. This plan was discarded after about a week. I am not too sure why but I just got into a rhythm of writing down the references manually in a word file. It seems to be harder to do it this way but for me it just worked.

I finished my draft literature review and submitted it. As this was the first time I had ever attempted anything to do with a research paper I really didn't know if the feedback would be good or if I would receive an audio file of someone breaking their heart laughing. Luckily it was the former. I had also been keeping all my files, documents, python code, references file in a folder on my desktop. This folder was synced with a repository on GitHub. This is an online change management system which is normally used for programming project but can hold any type of file. As I was used to this I decided to stick with what I knew.

Back to the research. As I had already tested the code to gather the twitter feeds I now just needed to get the list of athletes that I would analysing. I tracked down the twitter names of the top 10 male and female athletes based on the number of twitter followers. Then I put these names into my ode and harvested the data. I had found out in my testing that I couldn't harvest tweets from specific dates so I just harvested 5,000 tweets from each athlete and harvested the date it was posted. I then wrote some code to pick out the tweets from the correct time period.

The next thing I had to do was 'clean' the data. This was a bit of a mess at first as I had issues with ANSCII, UTF-8 and other character encodings. I had downloaded the data with one encoding and then I was writing my code to extract with different encodings. It took a bit of time before I realised my error and fixed the problem. When I sorted this out I then tried a few things to clean the data but one thing I realised quickly was that a lot of the tweets were not in English. I had a look at translating the relevant tweets into English but any translation will change the contents to some degree. I also thought about trying to analyse the non English tweets together as they were all Spanish but after looking in to this it was clear that this would not work. Another email to Mary Galvin and it was decided to only select athletes who tweeted in English and then reference this issue in the final research paper.

With these issues out of the way I now had 20 individual text files (one for each athlete) each containing cleaned, date specific tweets. I combined these into two large files (male and female). I then ran these through a cluster analysis code that I had written and tested at the start of the year. This produced the relevant number of clusters but the results of those clusters was less than inspiring. It was easy to spot that I had forgotten to put in a stop word list. Stop words are words like: the, and, my ,for, etc. The Python library I was using, Sklearn (this was used for the cluster analysis), had a stop word list built in. I used this and the results were better but I saw that I needed to add some more word to the list. This I did and I was left with my clusters of words.

At first glance it seemed that there was nothing really to the clusters and this was a bit disheartening. This of course was me looking at the clusters and when nothing immediately jump out at me thinking, 'it's all been a waste'. When I actually took a breath I realised that there was in fact something to the clusters. The clusters showed that there was a difference between male and female athletes in their self-presentation on Twitter.

Next I put the data files through a Python script that analysed the word counts. This confirmed the results of the cluster analysis.

Finally it came to writing up the dissertation. This was a daunting task. Thankfully Mary had sent those in her group an outline of how the dissertation should be done. This was a huge help. So following Marys guidance I wrote the dissertation. It did feel at times like it was some kind of 'brain dump' and I really struggled to get the paper to 'flow'. I consoled myself with the fact that at least I had the structure of the paper done and the main parts were in. I took a break for a few days and then went back to it with a fresh head. It wasn't a sooth process but after a good few rewrites I had something that I could send back to Mary and see what she thought.

Marys feedback was very encouraging. The paper was definitely on the correct path so now I just had to clean it up. It sounded easy but it took a good bit of work as I had taken two weeks off and getting back into it (remembering what way I was thinking) was not too easy. Another few days of rewriting, adding and deleting pieces and I finally finished my dissertation.

I am never getting in to the area of writing research papers!!