Capstone Project Findings Report

Anna Vaughan

January 16, 2018

```
## -- Attaching packages --------------------------------- tidyverse 1.2.1 --

## v ggplot2 2.2.1     v purrr   0.2.4
## v tibble  1.4.1     v dplyr   0.7.4
## v tidyr   0.7.2     v stringr 1.2.0
## v readr   1.1.1     v forcats 0.2.0

## -- Conflicts ------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

## Warning: Missing column names filled in: 'X4' [4]

## Parsed with column specification:
## cols(
##   `Cycle Year` = col_integer(),
##   `Applicant Domicile (UK / EU / Not EU)` = col_character(),
##   `Number of Acceptances` = col_integer(),
##   X4 = col_character()
## )

## Warning in rbind(names(probs), probs_f): number of columns of result is not
## a multiple of vector length (arg 1)

## Warning: 33 parsing failures.
## row # A tibble: 5 x 5 col     row col   expected  actual    file
## ... ................. ... ..........................................................................
## See problems(...) for more details.

## Warning: Missing column names filled in: 'X5' [5]

## Parsed with column specification:
## cols(
##   `Cycle Year` = col_integer(),
##   `Subject Group (Detailed Level)` = col_character(),
##   `Applicant Domicile (Region)` = col_character(),
##   `Number of Applications` = col_integer(),
##   X5 = col_character()
## )

## Warning in rbind(names(probs), probs_f): number of columns of result is not
## a multiple of vector length (arg 1)

## Warning: 20430 parsing failures.
## row # A tibble: 5 x 5 col     row col   expected  actual    file
## ... ................. ... ..........................................................................
## See problems(...) for more details.

## Warning: Missing column names filled in: 'X5' [5]

## Parsed with column specification:
## cols(
##   `Cycle Year` = col_integer(),
```

```
##   `Subject Group (Detailed Level)` = col_character(),
##   `Age Band` = col_character(),
##   `Number of Applications` = col_integer(),
##   X5 = col_character()
## )
```

```
## Warning in rbind(names(probs), probs_f): number of columns of result is not
## a multiple of vector length (arg 1)
```

```
## Warning: 8817 parsing failures.
## row # A tibble: 5 x 5 col     row col   expected  actual    file
## ... ................. ... ..........................................................................
## See problems(...) for more details.
```

```
## Warning: Missing column names filled in: 'X5' [5]
```

```
## Parsed with column specification:
## cols(
##   `Cycle Year` = col_integer(),
##   `Provider Name` = col_character(),
##   `Applicant Domicile (Region)` = col_character(),
##   `Number of Applications` = col_integer(),
##   X5 = col_character()
## )
```

```
## Warning in rbind(names(probs), probs_f): number of columns of result is not
## a multiple of vector length (arg 1)
```

```
## Warning: 30850 parsing failures.
## row # A tibble: 5 x 5 col     row col   expected  actual    file
## ... ................. ... ..........................................................................
## See problems(...) for more details.
```

```
## Classes 'grouped_df', 'tbl_df', 'tbl' and 'data.frame':  11 obs. of  3 variables:
##  $ cycle_year                                : int  2006 2007 2008 2009 2010 2011 2012 2013 2014 201
##  $ M40.The.Manchester.Metropolitan.University: int  39715 40860 38185 43535 50675 54965 45985 48555 5
##  $ S03.The.University.of.Salford             : int  16325 17610 17485 18095 20405 21810 18910 18660 1
##  - attr(*, "vars")= chr "cycle_year"
##  - attr(*, "drop")= logi TRUE
##  - attr(*, "indices")=List of 11
##   ..$ : int 0
##   ..$ : int 1
##   ..$ : int 2
##   ..$ : int 3
##   ..$ : int 4
##   ..$ : int 5
##   ..$ : int 6
##   ..$ : int 7
##   ..$ : int 8
##   ..$ : int 9
##   ..$ : int 10
##  - attr(*, "group_sizes")= int  1 1 1 1 1 1 1 1 1 1 ...
##  - attr(*, "biggest_group_size")= int 1
##  - attr(*, "labels")='data.frame':   11 obs. of  1 variable:
##   ..$ cycle_year: int  2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 ...
##   ..- attr(*, "vars")= chr "cycle_year"
##   ..- attr(*, "drop")= logi TRUE
```

```
## Warning: Missing column names filled in: 'X5' [5]
```

```
## Parsed with column specification:
## cols(
##   `Cycle Year` = col_integer(),
##   `Provider Name` = col_character(),
##   `Subject Group (Summary Level)` = col_character(),
##   `Number of Applications` = col_integer(),
##   X5 = col_character()
## )
```

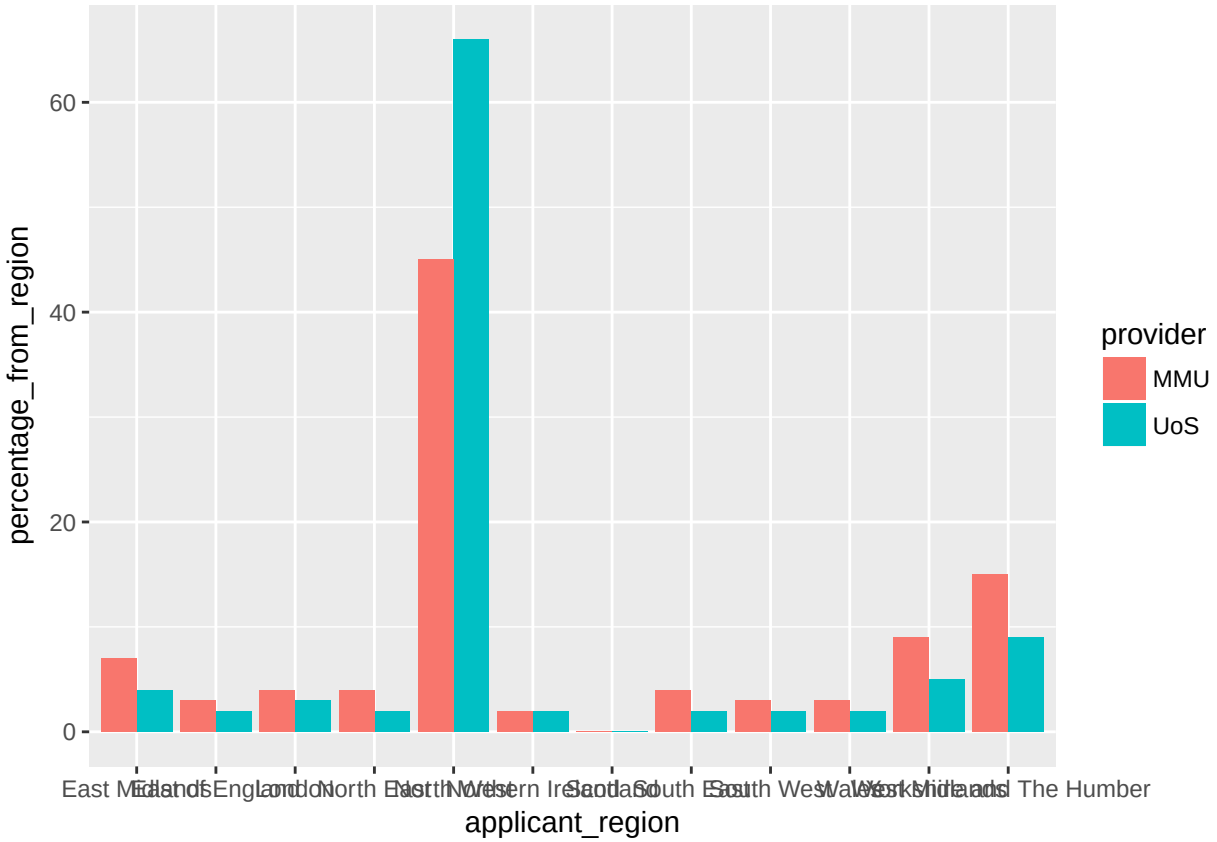
```
## Warning in rbind(names(probs), probs_f): number of columns of result is not
## a multiple of vector length (arg 1)
```

```
## Warning: 33814 parsing failures.
## row # A tibble: 5 x 5 col     row col   expected  actual    file
## ... ................. ... .....................................................................
## See problems(...) for more details.
```

```
## Warning in providername == c("The Manchester Metropolitan University", "The
## University of Salford"): longer object length is not a multiple of shorter
## object length
```

```
## Warning in providername == c("The Manchester Metropolitan University", "The
## University of Salford"): longer object length is not a multiple of shorter
## object length
```

```
## Joining, by = c("cycle_year", "provider_name", "applicant_domicile_(region)", "number_of_applications
```



Capstone Project Findings Report

A summary of the most interesting/important findings from my analysis

Revised problem that my Capstone project will solve

The overall aim is to understand where MMU's applications are stronger than UoS's (in terms of application numbers, subject balance and applicant regional origin), to help UoS academics and marketing teams understand where recruitment opportunities may be being missed, so that they can focus effort and resources on those opportunities. - How do overall UCAS application numbers to the University of Salford (UoS) compare with application numbers to Manchester Metropolitan University (MMU)? Who receives more applications? - How did overall UCAS application numbers to each university change over time from 2006 to 2016? Did one grow ahead of the other? - How does the spread of subject groups applied for at MMU and Salford compare? - How does the spread of subjects applied for UK-wide compare to the spread at MMU and UoS? - How does the spread of applicant domiciles at MMU compare with that at UoS?

Overall UCAS application numbers to UoS vs MMU

- What were application numbers at MMU and Salford, across the period 2006 to 2016? Reshaping the data in MMU_Salf_App_Numbers_4 showed the data just for the two institutions being compared: application numbers to MMU look to be consistently larger than those for Salford for each year, with application numbers to MMU mostly more than double those to Salford, though in some years the difference was larger than others.

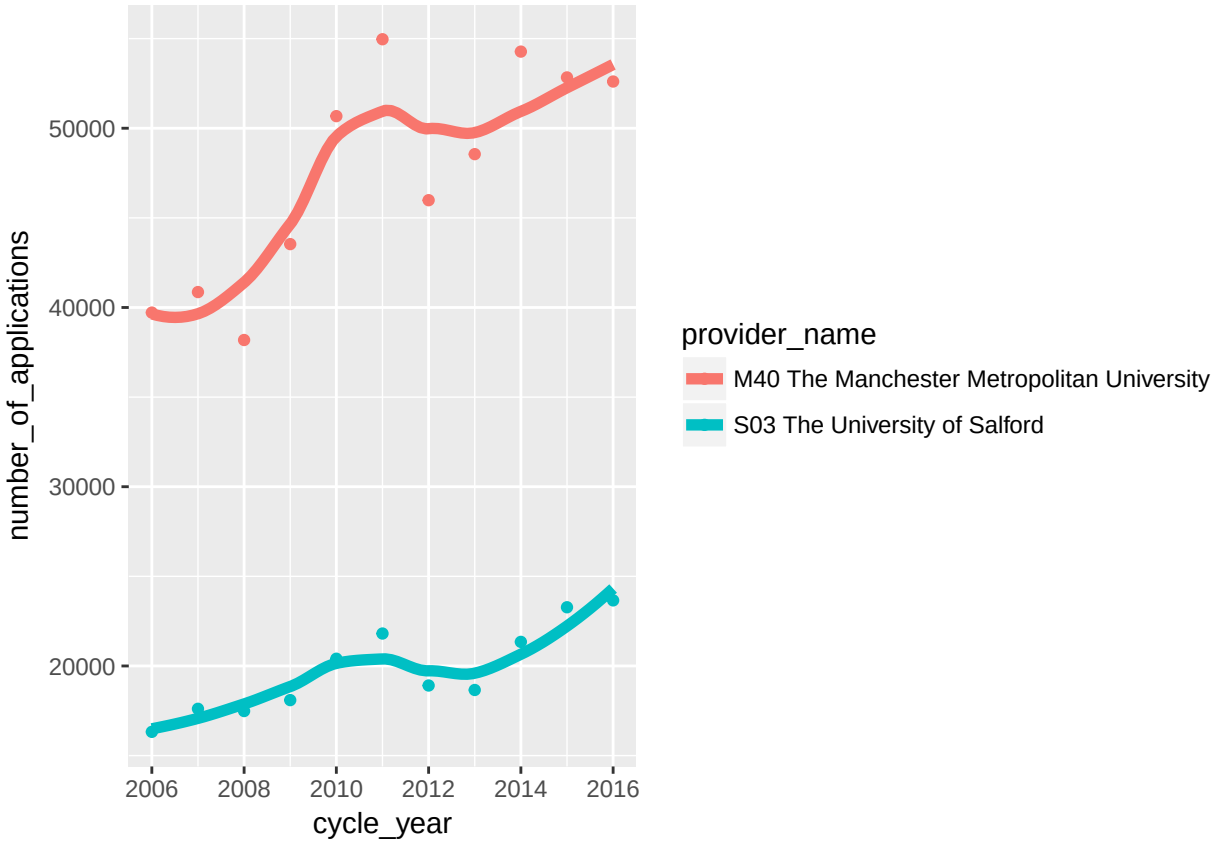
Reshaping the data to MMU_Salf_App_Numbers_6 showed the data as one institution per column just for the two institutions being compared, allowing me to sum applications across the 11 years: total application numbers to MMU were 522,190 (70.5% of the two institutions' combined total), while those to Salford were 217,580 (29.5% of the total).

INSERT TABLE FROM DATA FRAME MMU_Salf_App_Numbers_6

Trends in UCAS application numbers to UoS vs MMU over time

- How have application number changed over time at MMU and UoS? Plotting the application numbers to each instituion by year as a scatter plot and applying a trend line helps clarify trends in application numbers.

```
## `geom_smooth()` using method = 'loess'
```

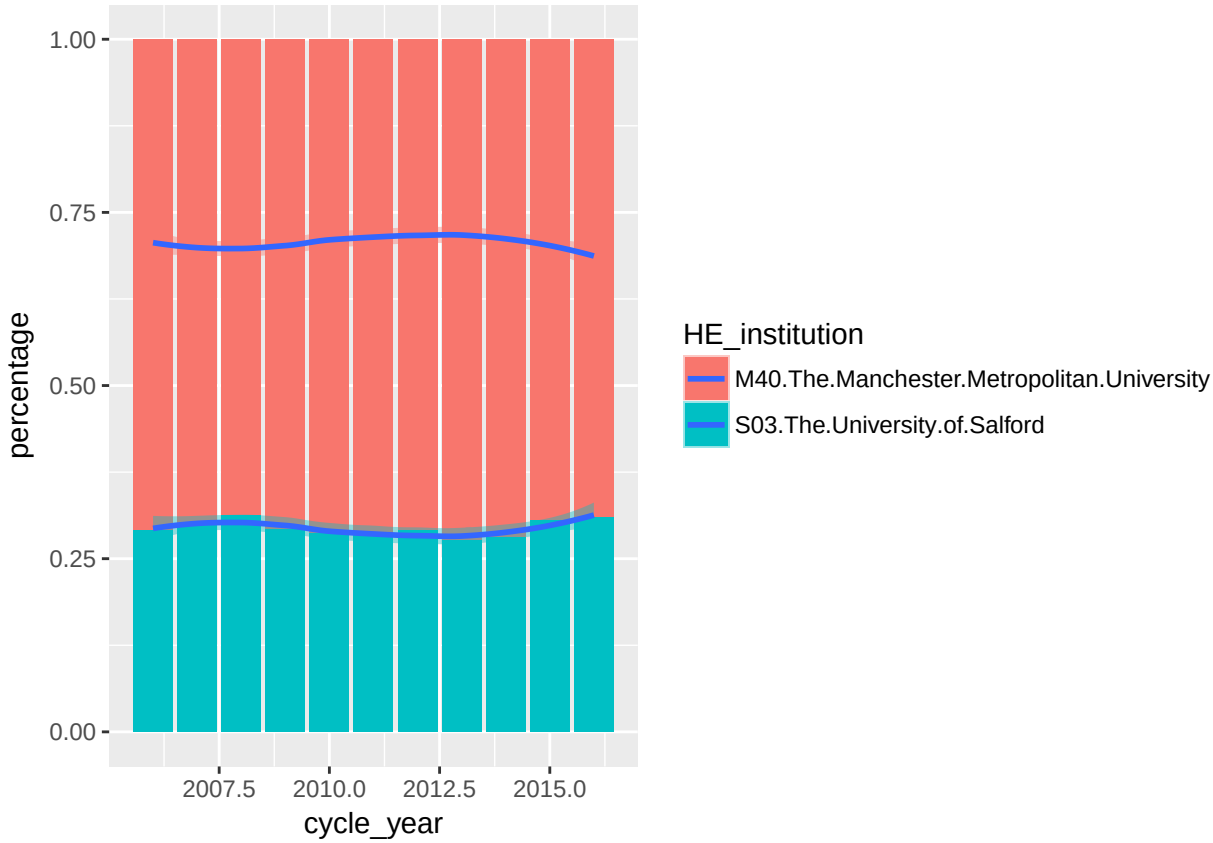


UoS saw a consistent increase in applications 2006 to 2011, then a sharp fall in 2012 followed by a gradual increase to 2016. UoS received its smallest number of applications in 2006 (16,325) and its largest in 2016 (23,660).

MMU's application numbers followed a more complex pattern: in addition to the pattern seen in UoS's numbers, MMU also saw a dip in applications in 2008, and saw fewer applications in 2015 and 2016 than it did in 2014.

Plotting a histogram showing the percentage share of the total applications to both institutions that each institution received helps in understanding the relative percentage of applications received by each institution regardless of the total application numbers.

```
## `geom_smooth()` using method = 'loess'
```



The plot shows that UoS's percentage share of the applications to both institutions fluctuated between 27% and 32% across the period 2006-2016, with UoS's share increasing fairly steadily in recent years (between 2013 and 2016).

(There seems to be far more variability in MMU's applications than Salford's year-on-year, but it would be good to look at the variance and standard deviation to check this).

How does the spread of subject groups applied for at MMU and Salford compare?

A table showing the percentage of an institutions' total applications that are for each subject groups helps clarify the balance of applications by subject group across the two institutions.

INSERT TABLE FROM DATA FRAME Subject_Group_2016_MMU_UoS_Percentage_Only_Spread_Totals

The table shows that 10% more of MMU's total applications are for Business and Admin Studies than of UoS's total applications (21% vs 11%), that 6% more are for Biological Sciences (12% vs 6%), and 3% more are for each of Social Studies (8% vs 5%), History and Philosophical studies (3% vs 0%), and Creative Arts and Design (14% vs 11%).

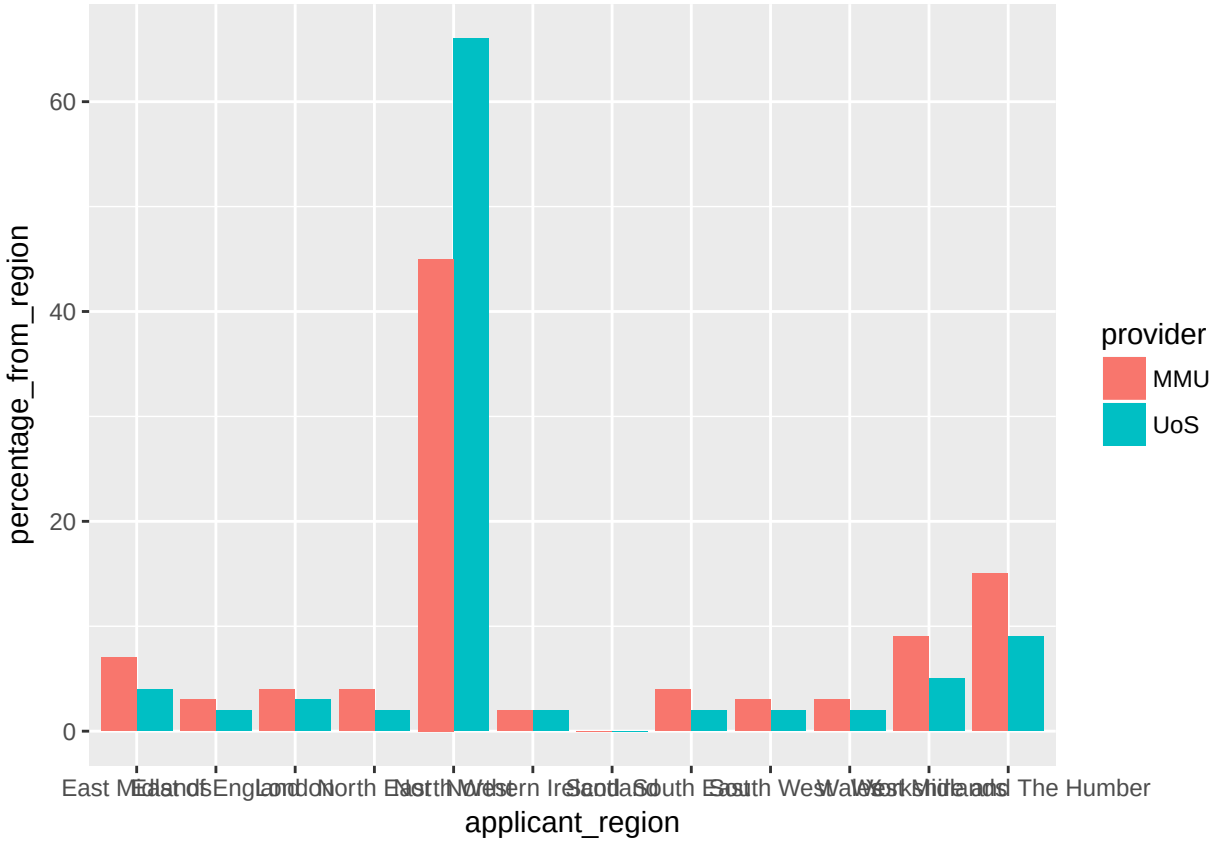
As these are subject areas where not only does MMU receive a greater volume of applications, but it receives a higher percentage of its total applications, these would be potentially fruitful areas for UoS's academics and marketing team to target to increase applications.

The data also suggests that each institution is heavily reliant on particular subject areas for a large proportion of its applications: MMU receives 21% of all its applications for the Business and Admin Studies subject area, while UoS receives 29% of all its applications for the Subjects Allied to Medicine subject area. While these subject areas are traditional strengths for each institution, it would be worth the marketing teams in particular being aware of the institutions' current heavy reliance on these subjects so that they could

diversify the application spread and reduce the risk to university income from a future downturn in their one key subject area.

How does the spread of applicant domiciles at MMU compare with that at UoS?

Creating a column chart showing the percentage of each of MMU's and UoS's applications from each UK region shows that UoS is more heavily reliant on applications from the local region (North West) than MMU, with two thirds of 2016 applications (66%) coming from this region for UoS, compared with only 45% for MMU.



The regions outside the North West that were providing the largest percentages of applications to MMU were Yorkshire and the Humber(15%), West Midlands (9%) and East Midlands (7%), suggesting that UoS should target these regions in particular to boost applications in the future.