R Markdown for Capstone - Data Import

Anna Vaughan

January 16, 2018

# Capstone Project - Data Import

## I wanted to import all the UCAS csv files I might need in one command, first creating a list of all the csv files with UCAS data in them, to help me choose which to work on

### First I checked and set my working directory

getwd() setwd(“C:/Users/AnnaVaughan/Capstone/Capstone”)

### I read a description of 1 file as a test

readLines((“UCASData/EOC\_data\_resource\_2016-DR2\_051\_06.csv”), n = 3)

### Then created a list of all files (reducing their filenames to not include the extensions, and removing all non-csv files)

allfiles <- list.files(path = “C:/Users/AnnaVaughan/Desktop/UCASData”) allfilenames <- substr(allfiles,1,33) allcsvfilenames <- allfilenames [1:285]

### I tried out reading a single csv file, importing it an creating a filename

testimport <- read\_csv(“UCASData/EOC\_data\_resource\_2016-DR2\_001\_01.csv”, col\_names = TRUE, skip = 5) names <-substr(filenames,1,33)EOC\_data\_resource\_2016\_DR3\_027\_03 <- read\_csv(“UCASData/EOC\_data\_resource\_2016-DR3\_027\_03.csv”, col\_names = TRUE, skip = 5)

## After trying for several hours using lapply and other functions, I decided instead to import the 3 files that I thought would be most useful in understanding the profile of applications to Salford and other institutions in actionable ways.

### These were the ones that included details on student domicile, higher education instituion applied for, and subject applied for. The three most useful files from this perspective seemed to be a) HE Provider by Student UK Region Domicile, b) Subject Applied For by Student UK Region Domicile, and c) Subject Applied for by Age.

EOC\_data\_resource\_2016\_DR3\_027\_03 <- read\_csv(“UCASData/EOC\_data\_resource\_2016-DR3\_027\_03.csv”, col\_names = TRUE, skip = 5) EOC\_data\_resource\_2016\_DR3\_018\_03 <- read\_csv(“UCASData/EOC\_data\_resource\_2016-DR3\_018\_03.csv”, col\_names = TRUE, skip = 5) EOC\_data\_resource\_2016\_DR4\_005\_03 <- read\_csv(“UCASData/EOC\_data\_resource\_2016-DR4\_005\_03.csv”, col\_names = TRUE, skip = 5)