Data engineering assignment

OULAD dataset code exports for the report

Kory

2024-03-24

## # A tibble: 1 × 1  
## n  
## <int>  
## 1 23369

## # A tibble: 1 × 1  
## n  
## <int>  
## 1 188

## # A tibble: 1 × 1  
## n  
## <int>  
## 1 6364

## [1] 25

## [1] 75.72417

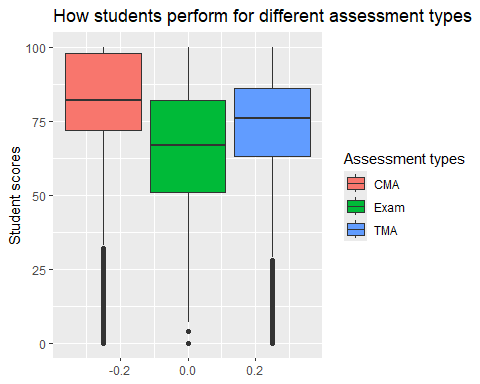
## [1] 80

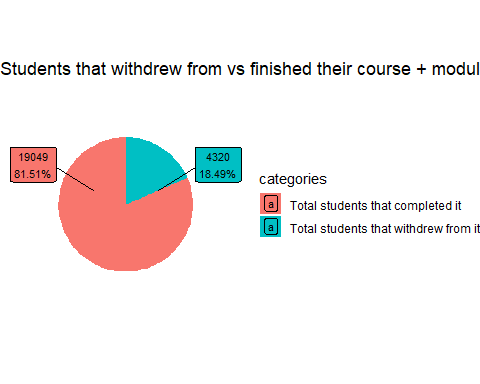
## 25%   
## 65

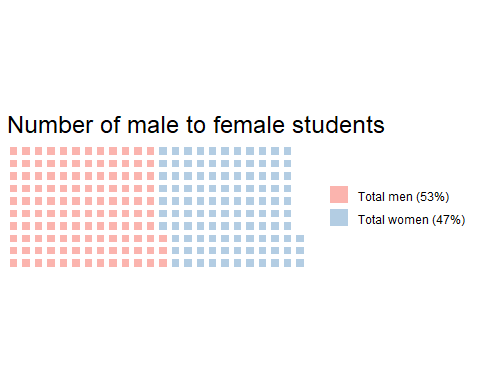
## 75%   
## 90

## [1] 0

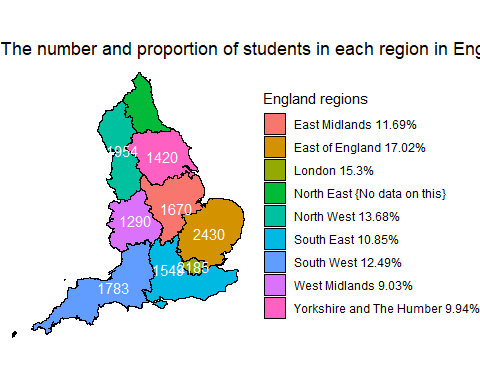
## [1] 100







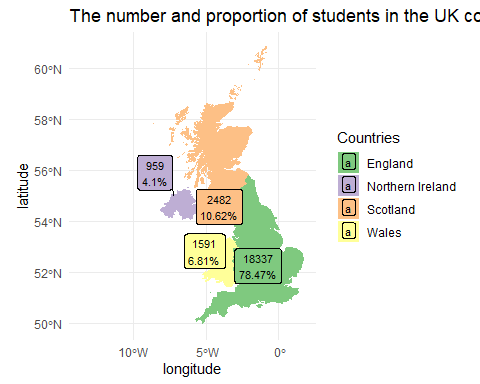
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom\_text()`).



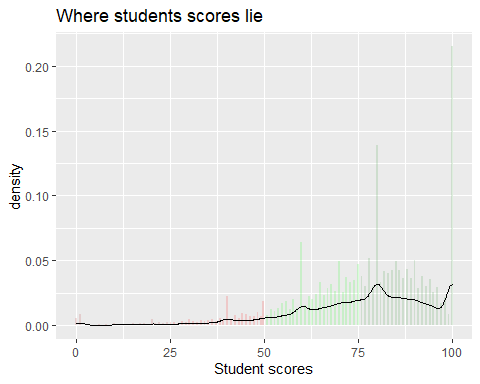
## Warning: Use of `uk\_countries\_with\_students\_count\_column$students\_in\_country` is  
## discouraged.  
## ℹ Use `students\_in\_country` instead.

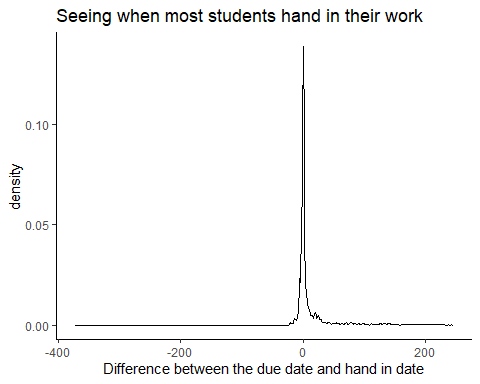
## Warning: Use of `uk\_countries\_with\_students\_count\_column$countriesTotal` is discouraged.  
## ℹ Use `countriesTotal` instead.

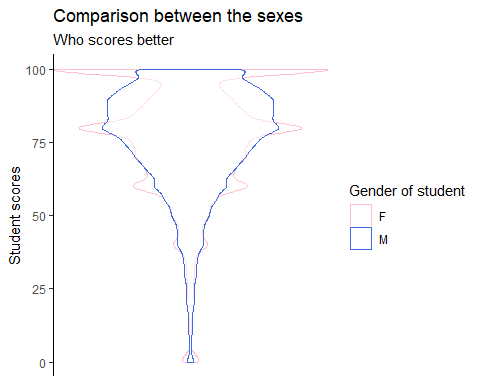
## Warning: Removed 228 rows containing missing values or values outside the scale range  
## (`geom\_label\_repel()`).



## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.  
## ℹ Please use `after\_stat(density)` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.

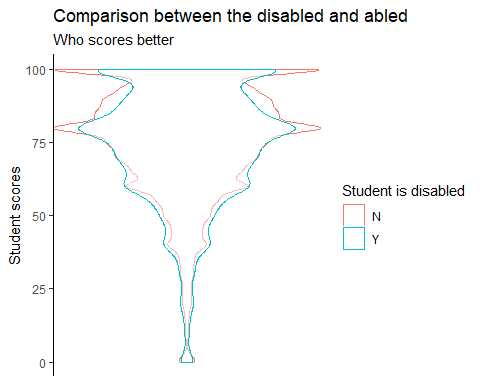




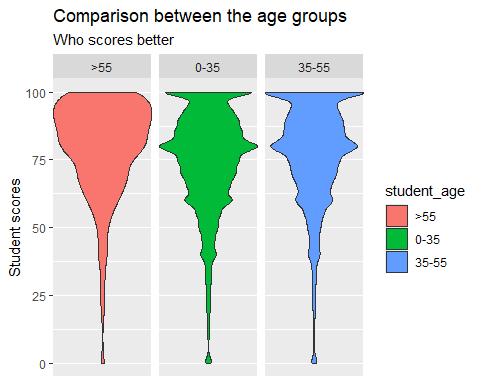


## # A tibble: 2 × 3  
## gender median mean  
## <chr> <dbl> <dbl>  
## 1 F 80 75.8  
## 2 M 80 75.6

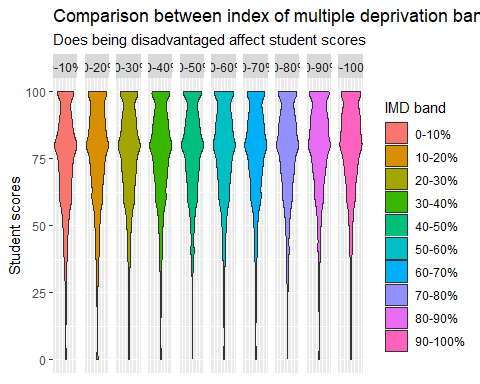
## # A tibble: 2 × 3  
## student\_has\_a\_disability median mean  
## <chr> <dbl> <dbl>  
## 1 N 80 75.9  
## 2 Y 77 73.5



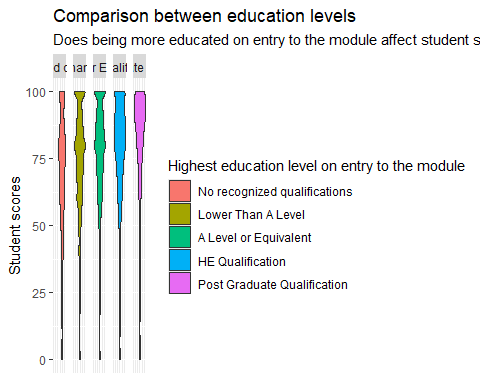
## # A tibble: 3 × 3  
## student\_age median mean  
## <chr> <dbl> <dbl>  
## 1 0-35 78 74.9  
## 2 35-55 80 77.5  
## 3 >55 83 79.4



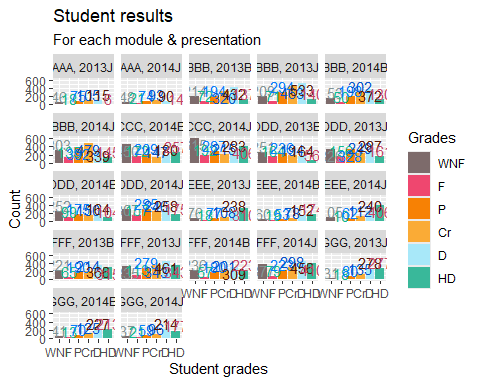
## # A tibble: 10 × 3  
## index\_of\_multiple\_deprivation\_for\_a\_uk\_region median mean  
## <chr> <dbl> <dbl>  
## 1 0-10% 77 73.1  
## 2 10-20% 79 75.1  
## 3 20-30% 78 74.8  
## 4 30-40% 79 75.2  
## 5 40-50% 80 75.8  
## 6 50-60% 80 75.8  
## 7 60-70% 80 76.2  
## 8 70-80% 80 76.4  
## 9 80-90% 80 77.7  
## 10 90-100% 80 78.0



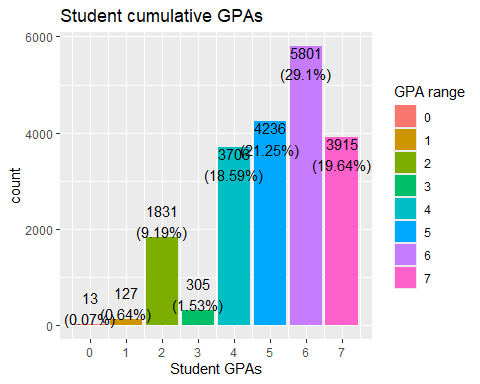
## # A tibble: 5 × 3  
## student\_highest\_education\_level\_on\_entry\_to\_the\_module median mean  
## <fct> <dbl> <dbl>  
## 1 No recognized qualifications 74 70.6  
## 2 Lower Than A Level 78 74.1  
## 3 A Level or Equivalent 80 76.2  
## 4 HE Qualification 81 77.8  
## 5 Post Graduate Qualification 88 83.5



## # A tibble: 22 × 3  
## group\_ids\_together median mean  
## <chr> <dbl> <dbl>  
## 1 AAA, 2013J 5 4.97  
## 2 AAA, 2014J 5 4.90  
## 3 BBB, 2013B 5 5.23  
## 4 BBB, 2013J 5 5.18  
## 5 BBB, 2014B 5 5.25  
## 6 BBB, 2014J 5 4.84  
## 7 CCC, 2014B 5 4.88  
## 8 CCC, 2014J 5 5.10  
## 9 DDD, 2013B 5 4.34  
## 10 DDD, 2013J 5 4.78  
## # ℹ 12 more rows



## Warning: Removed 3435 rows containing non-finite outside the scale range  
## (`stat\_count()`).  
## Removed 3435 rows containing non-finite outside the scale range  
## (`stat\_count()`).



## # A tibble: 1 × 1  
## n  
## <int>  
## 1 14020

## n  
## 1 59.99401

## # A tibble: 10 × 3  
## index\_of\_multiple\_deprivation\_fo…¹ total\_students\_that\_…² percentage\_of\_total  
## <chr> <int> <dbl>  
## 1 0-10% 537 21.0  
## 2 10-20% 559 18.2  
## 3 20-30% 543 21.1  
## 4 30-40% 461 17.5  
## 5 40-50% 445 18.3  
## 6 50-60% 406 16.8  
## 7 60-70% 409 18.3  
## 8 70-80% 359 16.1  
## 9 80-90% 340 15.7  
## 10 90-100% 309 15.4  
## # ℹ abbreviated names: ¹​index\_of\_multiple\_deprivation\_for\_a\_uk\_region,  
## # ²​total\_students\_that\_withdrew

## `geom\_smooth()` using formula = 'y ~ x'

