STAT 344 Group Project (Part I)

Lucas Qin, Frederick Wang, Isabel Wilson

2023-10-12

Reading in Data

```
all_data_files = list.files("data")
all_df = read.csv(paste0("data/", all_data_files[1]))

for (i in 2 : length(all_data_files)) {
  all_df = rbind(all_df, read.csv(paste0("data/", all_data_files[i])))
}

filtered_df = all_df %>% filter(Section == "OVERALL") %>%
  mutate(CourseNum = ifelse(
    is.na(Detail), Course, paste0(Course, Detail)
  )) %>%
  select(-Campus, -Year, -Session, -Section, -Professor, -Course, -Detail)

filtered_df %>% slice(1:10)
```

```
##    Subject                                                 Title Enrolled
## 1     AANB                              Topics in Animal Welfare       11
## 2     ACAM                        Asian Canadians in Popular Culture       50
## 3     ACAM                                Dis/Orienting Asian Canada       48
## 4     ACAM                           Selected Topics in ACAM Studies       53
## 5     ACAM                           Selected Topics in ACAM Studies       12
## 6     ACAM                      Asian Canadian Community-Based Media       18
## 7     ACAM Directed Studies in Asian Canadian and Asian Migration        9
## 8     ADHE                                          Teaching Adults      164
## 9     ADHE                           Institutions of Adult Education      142
## 10    ADHE       Developing Short Courses, Workshops and Seminars      166
##          Avg   Std.dev High Low X.50 X50.54 X55.59 X60.63 X64.67 X68.71 X72.75
## 1  92.72727  2.611165   95  90    0      0      0      0      0      0      0
## 2  77.96000 10.604119   93  46    1      2      1      1      2      2      5
## 3  77.60417 17.594414   91   5    3      1      0      0      1      1      3
## 4  80.24528 13.027937   92   8    1      1      1      0      0      4      5
## 5  80.91667  3.553701   86  75    0      0      0      0      0      0      2
## 6  84.94444  3.621378   90  77    0      0      0      0      0      0      0
## 7  94.33333  2.958040   98  90    0      0      0      0      0      0      0
## 8  81.92073 12.560521   97   0    2      2      3      0      4      9     10
## 9  84.06338 11.298161   98   0    2      0      0      7      0      2      9
## 10 83.60241 12.079474   98   0    2      0      0      1      6      5      8
```

```
##    X76.79 X80.84 X85.89 X90.100 CourseNum
## 1       0      0      0      11       550
## 2       3     17     12       4       250
## 3      10      8     16       5       300
## 4       2     14     18       7      320B
## 5       3      5      2       0      320D
## 6       1      6      9       2       350
## 7       0      0      0       9      447C
## 8      12     40     47      35       327
## 9       4     33     35      50       328
## 10     18     34     38      54       329
```

```
n = nrow(filtered_df)
n
```

```
## [1] 4263
```