Bài 4

Làm sạch data: Chuyển thời gian về dạng sử dụng được, mã số và điểm về dạng numeric. Loại bỏ NA .Tạo biến tgdata là khoảng thời gian dài nhất và landata là số lần nộp của từng trường hợp bằng group và ungroup.

1. Ta loại bỏ trong tgdata các thành phần lặp lại bằng [!dubplicate] và xuất ra adata
2. Ta sử dụng hàm count để tính dữ liệu phổ thời gian làm việc và sử dụng barplot để vẽ Barchart
3. Ta tính tần số bằng chia thời gian và số lần nộp.
4. Ta sử dụng filter để giữ lại những học sinh có tần số min và xuất ra những mã số min.
5. Ta sử dụng hàm count để tính dữ liệu phổ điểm, loại bỏ NA và dùng Barplot vẽ Barchart.
6. Ta sử dụng filter để giữ lại những học sinh có tần số max và xuất ra nrow() để tính số sinh viên.
7. In ra dữ liệu của câu f
8. Ta sử dụng hàm count để tính dữ liệu phổ điểm, loại bỏ NA và dùng Barplot vẽ Barchart.
9. Ta filter người có tần số cao nhất rồi lấy max để filter được người cao nhì, xuất dữ liệu ra
10. Ta tính y là tần số cao nhì của data và filter các học sinh có tần số cao hơn hoặc bằng a.
11. Ta sử dụng filter để chia data thành các hs nộp 1 lần và các hs nộp nhiều 1 lần:

* Với nộp 1 lần, ta cho thời gian TB bằng 0
* Với nhiều hơn 1 lần, ta lấy max thời gian của từng học sinh trừ đi các thời gian khác, lấy min để tìm ra thời gian TB lần kế cuối đến lần cuối (bỏ 0 để loại lần cuối trừ lần cuối)
* Dùng rbind để ghép 2 data frame

1. Ta sử dụng count để tính phổ trung bình thời gian, tính tần suất bằng cho n chia nrow, sử dụng vòng for để tính tần số tích lũy.
2. Ta sử dụng barplot để vẽ. Nhận xét: các biểu đồ có điểm chung là 0 sec chiếm đa số ( nếu không hầu hết dữ liệu )
3. Ta sử dụng pie để vẽ. Nhận xét: các biểu đồ có điểm chung là 0 sec chiếm đa số ( nếu không hầu hết dữ liệu )
4. Sử dụng barplot để vẽ. Nhận xét: các biểu đồ có điểm chung các tần suất tích lũy tang dần theo trung bình thời gian.
5. Ta sử dụng cấu trúc data %>% summarise\_all (‘Hàm cần tính’) để tính median, max, min.
6. Để đo mức độ phân tán xung quanh giá trị trung bình, ta tính khoảng biến thiên bằng lấy max – min, sử dụng cấu trúc data %>% summarise\_all (‘Hàm cần tính’) để tính var, sd (Phương sai và độ lệch chuẩn)
7. Ta sử dụng cấu trúc data %>% summarise\_all (‘Hàm cần tính’) để tính skrewness và kurtosis (sử dụng library(e1071)).
8. Ta sử dụng cấu trúc: quantile(data, c(0.25, 0.75), type = 1) để tính tứ phân vị 1 và 3