HW2 Instruction

This is a real job interview question from a data analysis company, and I doubt there is a standard answer to this question. So feel free to explore your story by using the data exploration and transformation techniques appropriately.

----------instruction quote begins-------------

Here is a small dataset for you to work with.

Each of 5 schools (A, B, C, D and E) is implementing the same math course this semester, with 35 lessons. There are 30 sections total. The semester is about 3/4 of the way through.

For each section, we record the number of students who are:

• very ahead (more than 5 lessons ahead)

• middling (5 lessons ahead to 0 lessons ahead)

• behind (1 to 5 lessons behind)

• more behind (6 to 10 lessons behind)

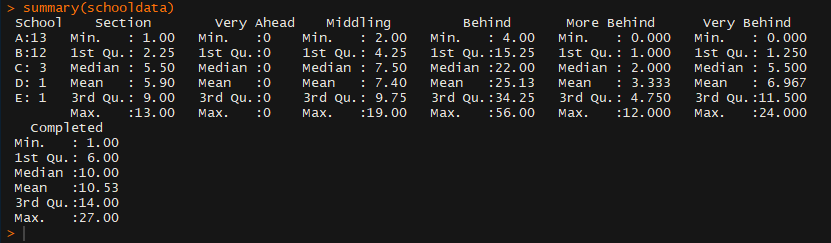
• very behind (more than 10 lessons behind)

• completed (finished with the course)

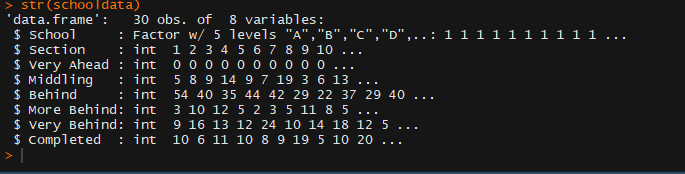
What’s the story (or stories) in this data? Find it, and tell it visually and, above all, truthfully.

-----------instruction quote ends-----------------

> summary(schooldata)



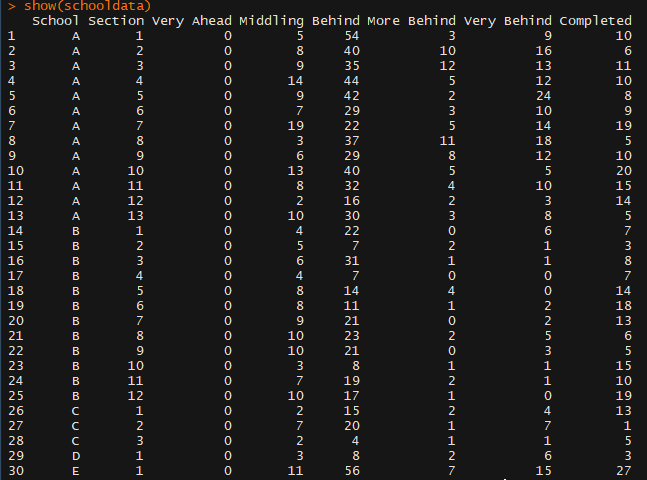
> str(schooldata)



Using **summary()** and **str()**, we can find out the summary and the structure of the given dataset

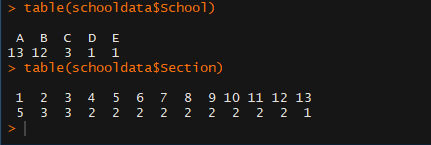
Here, we can see that all the variables are numeric, except for school.

> show(schooldata)



> table(schooldata$School)

> table(schooldata$Section)



Here we can see the number of math section across each school. School A has 13 sections, school B has 12 sections, school C has 3 sections, school D and E have 1 section each,

A screenshot of a cell phone

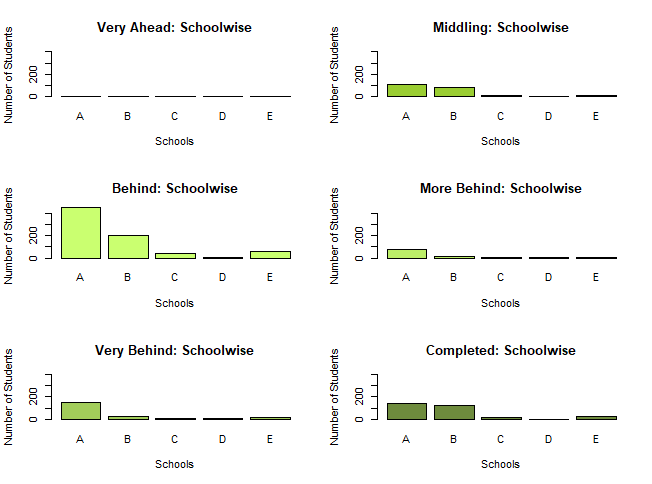
Description generated with high confidence

Here we can see that more students (approximately 750) are lagging by 1-5 lessons whereas number of students (approximately 100) lagging by 6-10 lessons is comparatively less. We can also see that no student is very ahead (by5 lessons).

A screenshot of a cell phone

Description generated with very high confidence

Here, we can see the same thing as earlier. More number of students are in range of behind by 1-5 lessons.



No school has students who are “Very ahead”

A is ahead of other schools in almost every other category, probably because **it has comparatively more students**.

Number of students in school C, E and D are very less.

**Conclusions:**

1. The number of students across schools are:

A > B > C > E > D

1. The number of students who are behind (1-5 lessons behind) are substantially higher than the other categories.
2. There are no students who are very ahead in the course, in any school.
3. Most of the students are 1-5 lessons behind, across all the schools.
4. Following is the order of sizes of levels of completion:
5. Behind > Completed > Middling > Very Behind > More Behind > Very Ahead