

The Data

The data set contains standardized Learning Optimization Koefficient (LOK) test scores for 1,000 students in Zachariatown during the 2019-2020 school year.

Columns

1. gender
2. race_ethnicity - race group that the student identifies with based on demographic survey filled by parent (the actual race is concealed for privacy purposes)
3. parental_level_of_education - the highest level of education completed by either parent/guardian
4. lunch - free/reduced OR standard
5. test_preparation_course - whether or not the student has completed the test preparation course
6. math_score - normalized score the student received on the math section of the LOK test
7. reading_score - normalized score the student received on the reading section of the LOK test
8. writing_score - normalized score the student received on the writing section of the LOK test

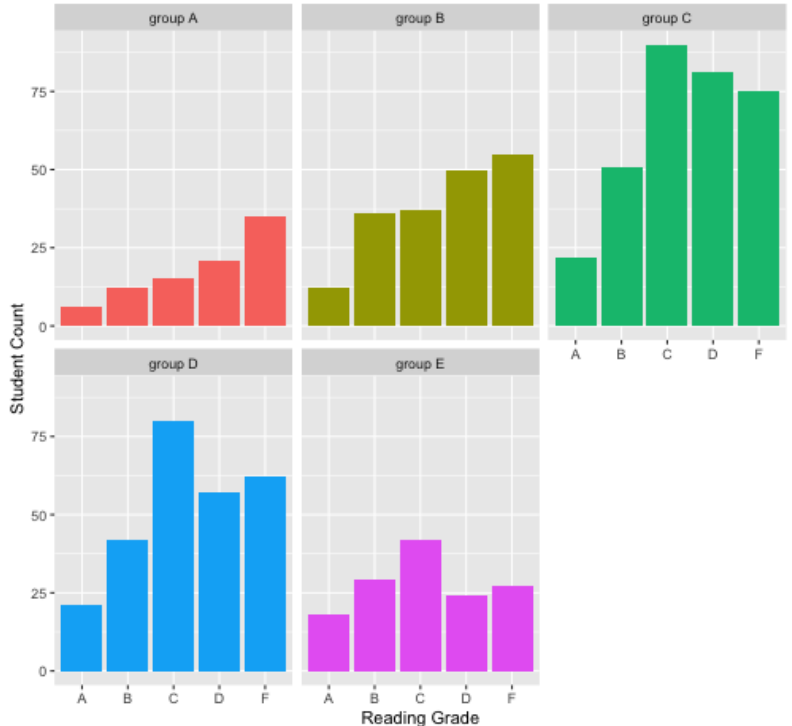
Questions

1. Load the data into your R session and set the data frame variable name as "perf"
 - a. The column `parental_level_of_education` should be defined as a factor with levels represented by the level of education. Change the column's data type so that it has the following levels (in order):
 - i. some high school
 - ii. high school
 - iii. some college
 - iv. associate's degree
 - v. bachelor's degree
 - vi. master's degree
 - b. Let's transform new columns to indicate a categorical letter grade for each of the test score values. These new columns should be named as {section}_grade (ex. math_score should have a corresponding math_grade; writing_score should have a corresponding writing_grade).
 - i. A: ≥ 90
 - ii. B: $80 \leq \text{score} < 90$

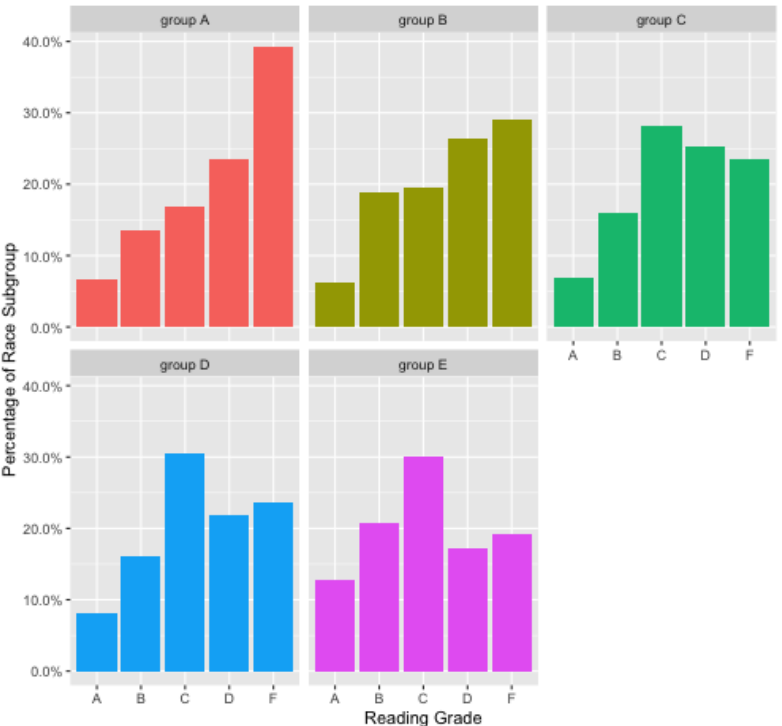
- iii. C: $70 \leq \text{score} < 80$
- iv. D: $60 \leq \text{score} < 70$
- v. F: < 60

2. Use the correct plot to visualize the relationship between a students' math and reading test scores.
 - a. Let's get a different level of granularity here. Is there a change in the relationship based on the parental level of education?
3. There is an assumption that female students perform better at comprehension questions than males, whereas male students are slightly more analytical. Think through some visualizations that could show evidence either supporting/rejecting this assumption.
4. The President of the Board of Education (BOE) of Zachariatown would like to assess whether there is a relationship between a student's race/ethnicity and his/her reading comprehension. Several researchers in the town have compiled different visualizations to attempt to answer this question, but the president has narrowed it down to four (seen on the next page).
 - a. Explain in 2-4 sentences what each plot is representing (eg. What does the x-axis represent? What are the units? What type of plot is being used?)
 - b. Which plot would you recommend to the BOE? If none, how would you approach the problem differently?

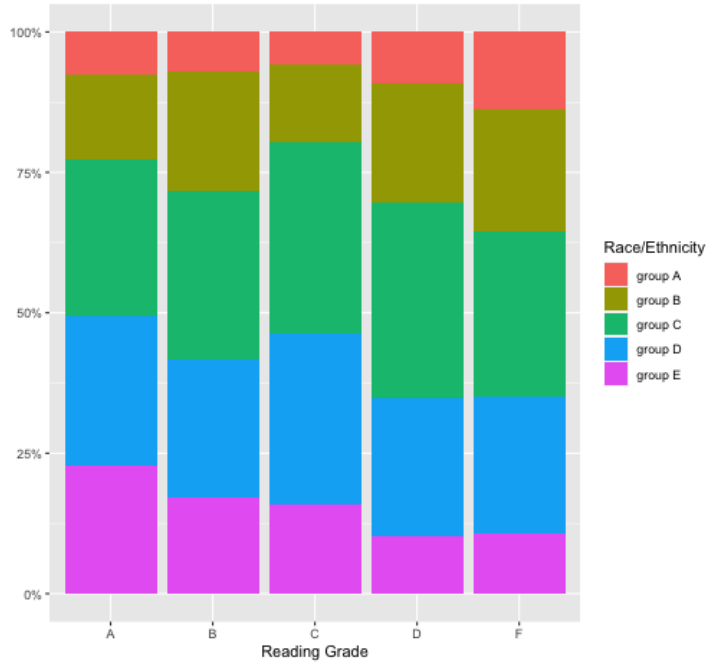
Distribution of Reading Grades by Race/Ethnicity
Learning Optimization Koefficient (2019-20)



Distribution of Reading Grades by Race/Ethnicity
Learning Optimization Koefficient (2019-20)



Distribution of Reading Grades by Race/Ethnicity
Learning Optimization Koefficient (2019-20)



Distribution of Reading Grades by Race/Ethnicity
Learning Optimization Koefficient (2019-20)

