Chapter 1

Exploration 1.1

1. Does the type of sequence affect memory?
2. *Response Variable* – number of letters recalled before mistake
   Variable Type – quantitative variable
   Unit – number of letters
3. Experiment, each student participant was randomly assigned to one of two sequences.
   The following answers use the example data in the MemorizingLettersCP.txt file from students at Cal Poly on the ISI-2 page where meaningful sequences are always in 3-letter groups (e.g., JFK-CIA-FBI-USA) and meaningless sequences are not always in 3-letter groups (e.g., JFKC-IAF-BIUS).
4. The 26 students are the 26 experimental units
5. *Treatment variable* – type of sequence
   Levels – meaningful, meaningless
6. Randomly assign the letter sequence to each student, this allows for confounding variables to be considered eliminated and causation to be inferred
7. Selected students from the same class, meaning they have followed similar educational paths and are of similar ages. We also didn't tell the students they were in different groups, and we made sure all of the testing conditions (e.g., instructions given, timing) were identical for both groups.
8. Double blind study, neither the student participants nor the teacher administrating the experiment were aware of which sequence was given to each student.
9. No, students were not randomly selected making this a convenience sample. In terms of memorization ability, these students may be similar to other college students their age, at universities similar to Cal Poly. One possible population is all college students similar to these at universities like Cal Poly.
10. In this particular dataset, we don't see evidence of incorrect observations and we don't know of any individuals with learning disorders or memory loss.
11. A large portion of the scores was between 15-20 letters, and the scores are relatively bell-shaped and symmetric. The mean score is 13.88 letters and the standard deviation of the scores is 6.62 letters.
12. *predicted score* = 13.88 letters
    SE of residuals = 6.62 letters
13. The meaningful sequence group tended to score higher by 3.77 letters (mean 12.00 vs. 15.77 letters); however, there was also more variation in the meaningful sequence group (SD 7.72 vs. 4.92 letters). The meaningful sequence group had all but two of their scores divisible by 3 which makes sense as the sequences were grouped into sets of 3 letters.
14. $\textit{predicted score} = \begin{cases} 15.77 \textit{ if meaningful} \\ 12.00 \textit{ if meaningless} \end{cases}$
    SE of residuals = 6.469 letters

15. The standard error of residuals is a little bit smaller, but the reduction amounts to less than half a letter.
16. No, there is still variation in the scores within each treatment group, because the group standard deviations are not 0. (Note: If the group standard deviations are not 0, then the SE of residuals won't be 0.)
17.

Observed Variation in: Memory score	**Sources of explained variation**	**Sources of unexplained variation**
Inclusion criteria: Student in class *Design:* Students started and stopped memorizing at the same time	Letter grouping sequence: meaningful vs. meaningless	Memorization ability, caffeine consumption, sleep, etc.

18. No, the difference in sleep means is only 0.19 hours or about 11 minutes (7.11 vs. 6.92 hours, for meaningless and meaningful, respectively).
19. Because caffeine is a categorical variable, the graph is two bar charts. A 0 represents "no caffeine" and a 1 represents "yes caffeine." The means of meaningless and meaningful are computed by adding up all of the 0's and 1's for each sequence, then dividing by the number of students in that group. Thus, the means are the proportion of students in that group who consumed caffeine prior to the study. For the meaningless group 69.2% of students consumed caffeine, whereas in the meaningful sequence group 53.8% consumed caffeine. Because there were 13 students in each group, this difference is about 2 students. With only a two-student difference, there probably is not enough of a difference to think that caffeine use is a confounding variable in the relationship between sequence and score.
20. Yes, but it is unlikely as random assignment guards against confounding variables.
21. We recommend the letter grouping model as there is more variation explained and the standard error of residuals is reduced from 6.623 to 6.469 letters. The difference in means is less than 4 letters, as this is a little less than 1 SE of the residuals, the difference does not appear to be all that meaningful.
    We would be willing to generalize conclusions from this study to college students similar to those who participated in the study at universities similar to Cal Poly. Cause-and-effect is valid because the study used random assignment, but only if the difference in means turns out to be statistically significant.
22. Conduct this test on a variety of classes or ages, make the sequences more distinct, etc.