1. The main program of the hash table project collects statistics from the search test. Use the program as an experiment to investigate the performance of your hash table. Write a paper discussing your results. Use the same format for your paper as described in Assignment 9. Your paper should address the following questions:

(a) Which is the better hash function, Knuth or Bernstein? Is one consistently better, or can you find conditions where Knuth is better than Bernstein and other conditions where Bernstein is better than Knuth?

(b) The text says that the divisor (number of slots in the hash table) should be a prime number and that particularly bad performance can occur if the divisor is a power of two. Can you find some nonprime numbers, specifically powers of two, that give particularly bad performance? It would be especially interesting to find a prime divisor that gives particularly good performance and a divisor that differs from it by only one that gives particularly poor performance. See this article on Mersenne primes for some ideas.

https://en.wikipedia.org/wiki/Mersenne_prime

Is one of the hash functions better at hashing with hash tables having the number of slots a power of two?

Based on your results from (a) and (b), use only the better hash function for your answers to the following questions.

(c) Is it possible to have approximately the same load factor $\alpha$ with hash tables having different numbers of slots? Would two such tables have about the same performance?

(d) How does the number of comparisons vary with the load factor in a table with a prime number of slots? Show the relationship with a graph including a wide range of $\alpha$ (from much less that 1, which would be typical, to greater than 1, which would not be typical). From your data, what value of alpha would you recommend for a good balance between storage required and speed of access?

Each of these questions should be answered by showing data that you collected from your experiment to address the question. Show your data in both graphical and tabular form.

Make maximum use of the many symbols in words.txt. That is, do not investigate small hash tables when you can investigate large ones that would be more realistic in practice.

Your paper must discuss the results of your experiment according to the following outline:

Title, Name, Date

Abstract
In one short paragraph, describe the purpose of your paper and your conclusions in general.

1. Introduction
Explain in one paragraph the assignment as if the reader were a stranger who does not know you or Pepperdine University. Explain in a second paragraph what each following section describes.

2. Method
Include the following subsections.

2.1 Hash tables
Describe the characteristics of a hash table. How does it work? What are collisions? How are they resolved? Define the load factor $\alpha$.

2.2 Hash functions
Describe the characteristics of a good hash function. Describe the Knuth and Berstein hash functions.

2.3 Data collection
Describe the OO design pattern used to take the data. Describe the computer runs that took the data.
2.4 Analysis
Explain the data that was collected by the test program. Include the results from a sample run and explain each line of the output.

3. Results
Include the following subsections with appropriate plots and/or tables.

3.1 Evaluation of hash functions
Answer the question about the Knuth vs Berstein hash function with data to justify your conclusion.

3.2 Hash table size
Answer the question about the effect of hash table size on performance.

3.3 Equal load factors
Answer the question about performance with equal load factors.

3.4 Effect of load factor on performance
Answer the question about performance as a function of load factor.

IMPORTANT: The writing style in this section should be to make a statement, then back it up by quoting data or referring to graphs, then discuss. Make another statement, back it up by quoting data or referring to graphs, then discuss, etc. If any data appears unusual or is not what you would expect, try to explain it. You might try analyzing the data a different way as a result of your analysis. If so, describe what you tried.

4. Conclusion
One paragraph of the conclusions from your experiment. This section is a summary of your results section with specific conclusions. It differs from the abstract, which summarizes the conclusions in general.

You must typeset your paper with LaTeX with the paper template sample you used in Assignment 9.

This assignment is to begin working on your paper. The final paper is due with Assignment 23. Name your document 99a23written.pdf where 99 is your assigned two-digit number.