

Each student has been assigned an English text selection to analyze (instructions for getting your assigned text are below). For this exam we are concerned with word length and letter frequency.

For your individual text selection, do the following:

1. Make a histogram of the word lengths (use a bin width of 1).
2. Compute the five-number summary for word lengths.
3. Make a box plot of word lengths.
4. Compute the mean, variance, and standard deviation of the word lengths.
5. Make a bar chart of the letter occurrences, i.e. how many times each letter occurs. Order the values by frequency, highest to lowest.

You have also been assigned a city. Go to the website [www.cityrating.com](http://www.cityrating.com) click on [Weather History](#) and find the weather history for your assigned city.

6. Make a time plot of the average monthly temperature for your assigned city.

How to get your specific data:

Go to the [www.staley-classes.org](http://www.staley-classes.org) website. In the Statistics column there should be a link called [Exam 1 Do-At-Home Data](#). Open that page and then next to your student ID number you will find a hyperlink for your text and a city name. Use the text link for 1-5 above. Count apostrophes as one letter, hyphens as zero letters, and ignore numbers. Use the city link for #6 above.

Recommendations:

Using Excel for this project is optional. If you choose to use Excel there are several **helpful movies** on the [www.staley-classes.org](http://www.staley-classes.org) website on pages:

[Chapter 1 Stuff](http://www.mr-ideahamster.com/classes/stats/M119%20C01.htm) (<http://www.mr-ideahamster.com/classes/stats/M119%20C01.htm> )

and

[Chapter 2 Stuff](http://www.mr-ideahamster.com/classes/stats/M119%20C02.htm) (<http://www.mr-ideahamster.com/classes/stats/M119%20C02.htm> )

and for working with text in Excel:

[Excel Suggestions](http://www.mr-ideahamster.com/classes/stats/movies/textanalysis_1-excel/textanalysis_1-excel.html) ([http://www.mr-ideahamster.com/classes/stats/movies/textanalysis\\_1-excel/textanalysis\\_1-excel.html](http://www.mr-ideahamster.com/classes/stats/movies/textanalysis_1-excel/textanalysis_1-excel.html) ).

Part of the scoring will be based on neatness, comprehensibility, quality of your labeling, etc.

**A Note about Scoring.** Here is a list of things that students in past semesters have done that have adversely affected their score on this DAH exam:

- a. They didn't label/title their graphs/tables.
- b. They didn't put the city of interest in the graph/table title for #6.
- c. They didn't capitalize the first letter of each word in the graph/table title.
- d. They used the frequency of word lengths as the data set rather than the word lengths themselves as the data.
- e. They had misspellings.
- f. They had some ridiculously long word lengths.
- g. They made their answers hard to find.
- h. Their graph in #5 had an incorrect ordering for the horizontal axis.
- i. They did not check their data for correctness. Being careful is an important part of this class.
- j. They divined words of length zero.