

# Statistics for Engineering, 4C3/6C3

## Assignment 1

Kevin Dunn, kevin.dunn@mcmaster.ca

Due date: 16 January 2014

---

**Assignment objectives: creating and interpreting data visualizations**

### Question 1 [3]

Which types of features can the human eye easily pick out of a time series plot?

### Question 2 [4]

*Final exam, 2013:* Why is the principle of minimizing “data ink” so important in an effective visualization? Give an engineering example of why this is important.

### Question 3 [10]

Reproduce the box plot for board thickness that was discussed in class. The board thickness data set is available from [the dataset website](#).

1. Reproduce the figure that was shown in class, using the first 100 rows from the data set. See R code in the course notes.
2. Create a new box plot using rows 3100 to 3300. Interpret any interesting observations from this box plot. Superimpose a target line of 1680 mils.
3. Explain why the thick center line in the box plot is not symmetrical with the outer edges of the box.

This question is to ensure you can install R and use the course dataset site.

### Question 4 [5]

Describe what the main difference(s) between a bar chart and a histogram are.

### Question 5 [8]

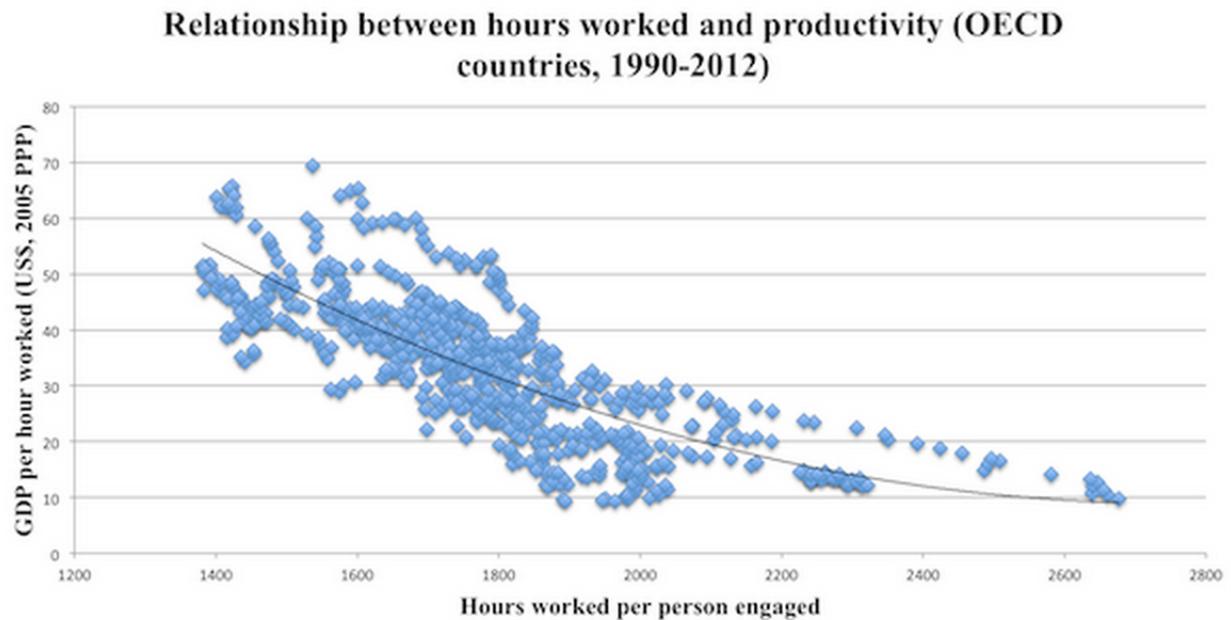
In a question on the final exam in 4M3 there was an open-ended question. The [data values are the grades](#) achieved for the answer to that question, broken down by whether the student used a systematic method, or not. No grades were given for using a systematic method; grades were awarded only on answering the question.

A systematic method is any method that assists the student with problem solving (e.g. define the problem, identify knowns/unknowns and assumptions, explore alternatives, plan a strategy, implement the strategy and then check the solution).

Draw two box plots next to each other that compare the two data sets. Also comment on any features you notice in the comparison.

### Question 6 [8]

Consider this plot we saw in class (it is also available on-line, with [some additional context](#))



1. What is the plot's author trying to convey with this scatter plot?
2. Do you believe this an effective and complete message (i.e. could you improve it somehow?)
3. Is there a causal mechanism at play between the two variables?
4. How would you confirm or disprove the message the plot's author is making?

### Question 7 [10]

At the start of the class several people indicated they wanted to learn about visualizing more than 3 variables. In class we saw a way to visualize at least 5 variables.

Here's another method that you can investigate. Read up about scatterplot matrices, and draw one for the [Food texture data set](#). See the `car` library in R to create an effective scatterplot matrix with the `scatterplotMatrix` function.

Give a couple of bullet-points interpreting the plot.

### Question 8 [0]

Read the short, clearly written article by Stephen Few on the pitfalls of pie charts: [Save the pies for dessert, http://www.perceptualedge.com/articles/08-21-07.pdf](http://www.perceptualedge.com/articles/08-21-07.pdf).

I do recommend you read this. The article presents an easy-to-read argument against pie charts that will hopefully convince you.

Here's a [great example](#) from the CRA.

---

END