

# Statistics for Engineering, 4C3/6C3

## Assignment 1

Kevin Dunn, kevin.dunn@mcmaster.ca

Due date: 18 January 2013

**Assignment objectives: create suitable data visualizations**

### Question 1 [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 4800 to 4900. 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.

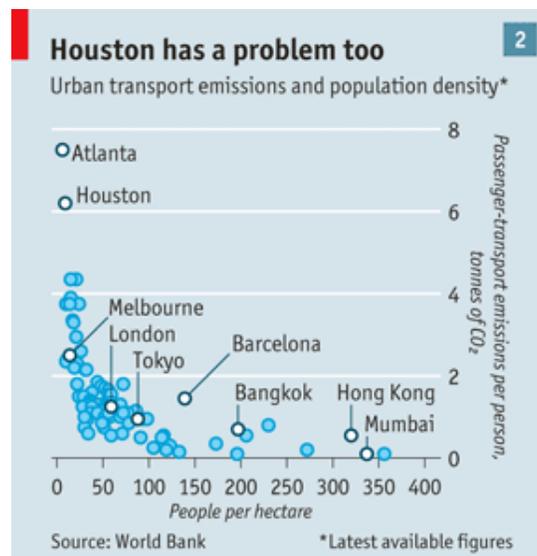
This question is to ensure you can install R and use the course dataset site.

### Question 2 [5]

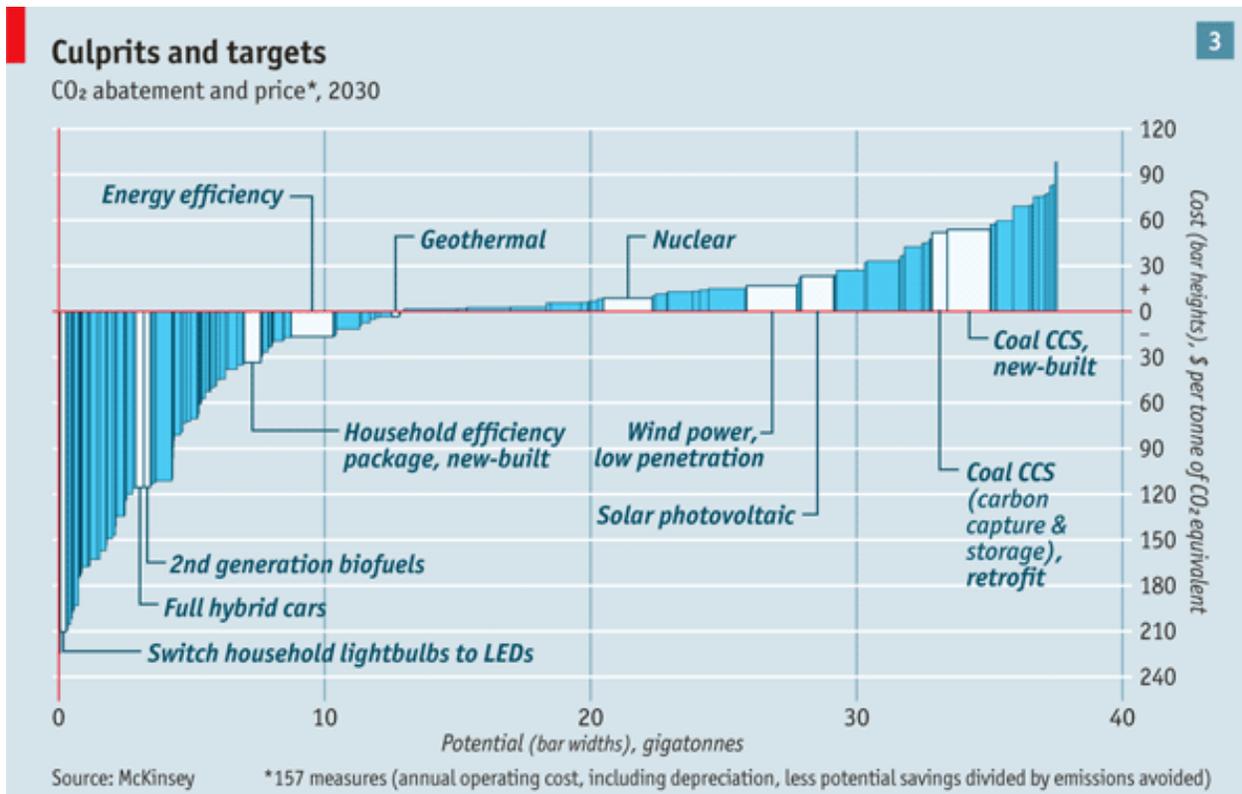
Describe what the main difference(s) between a bar chart and a histogram are.

### Question 3 [5]

In an article published in June 2012, what might this plot's author be asking you to infer?



Question 4 [6C3 students: 8]



Related to the previous question, a cost curve [created by a business consultancy](#), shows different options to cut carbon emission amounts ( $x$ -axis).

Describe whether this is an effective visualization. In your answer highlight some of the interesting information you are able to decode from the visual data.

Question 5 [5]

The instructor uses an app to track his GPS coordinates as he drives to work and back to Hamilton each day. The app collects the location and elevation data every 5 meters, or every 2 seconds, roughly 4000 data points per trip. Data for these trips are described [on the course website](#).

Plot an interesting visualization from these data. The visualization should be accompanied only by (a) the question you are trying to ask the data set, (b) the plot that you draw, and (c) a single short sentence that summarizes the answer to your question. In other words, the visualization should answer the question, not your written text.

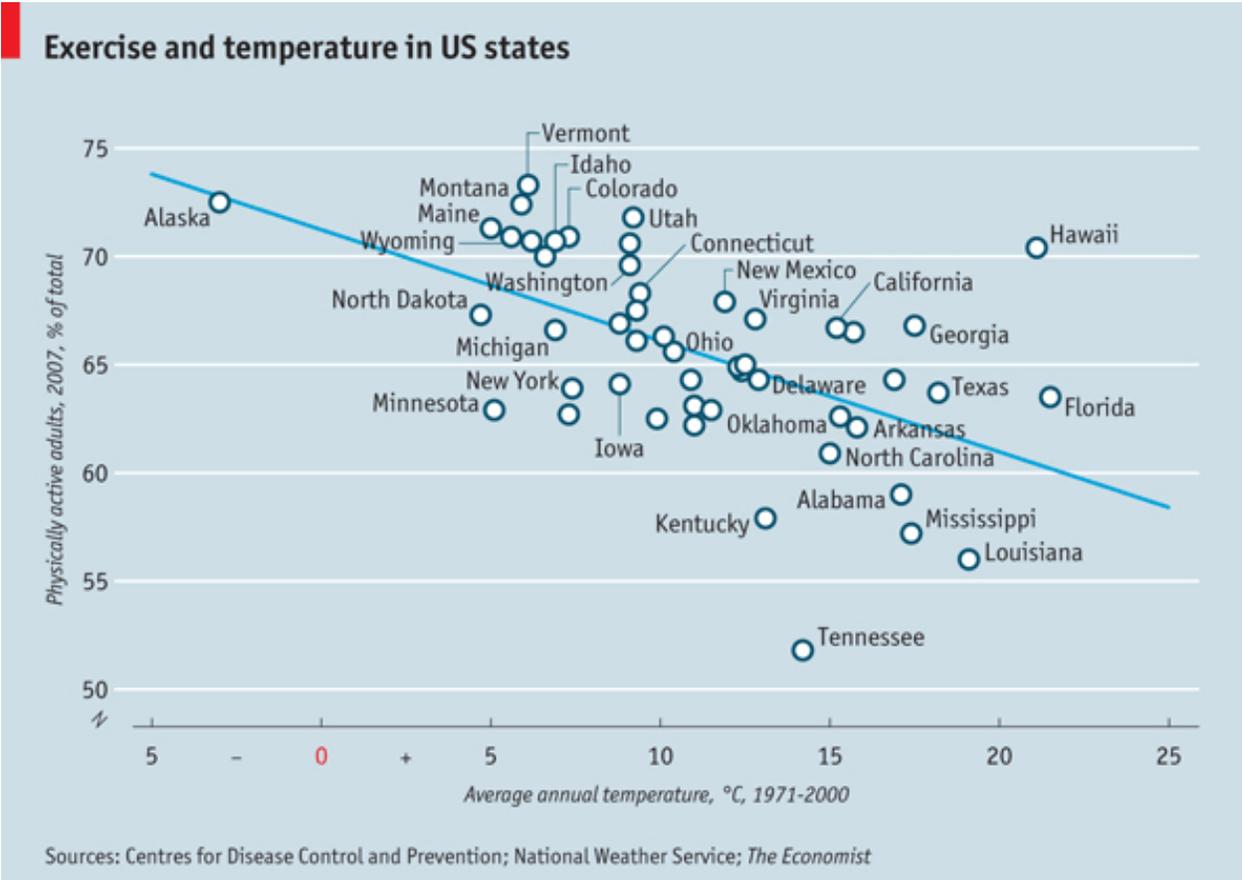
Questions should be interesting (i.e. not something like “what is the average trip duration”), but more challenging.

Please feel free to use R, Excel, MATLAB, Python, or any other tool to answer this question, but ensure your plots obey the general guidelines for excellent data graphics covered in this section of the course.

Question 6 [10]

From the 4C3/6C3 final exam, 2012 [8 out of 100 in the exam]

The following figure, taken from [The Economist](#) shows the percentage of physically active adults against the average annual temperature, broken down by geographical regions, according to the USA state.



1. Since visualization plots can often stand alone without accompanying text, what is the plot's author asking you to infer from this visualization? [2]
2. Is there a causal relationship in the data? Explain your answer. [2]
3. The author has shown a linear regression line. Is the intercept term meaningful in this case; please explain. [2]
4. Calculate an estimate of the linear model's slope, and give an interpretation for it. [2]

**Question 7 [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).

END