

36-315: Statistical Graphics and Visualization

Homework 1

Date: January 13, 2002

Due: start of class January 20, 2002

1. In class we discussed a graphical train schedule developed by E.J. Marey.
 - (a) Name two things that you can read from this plot that are not on an ordinary tabular schedule.
 - (b) Can you find a reason why, after 100 years, Marey's idea is not used in public transportation? In what ways is a table better?
2. In class we discussed the data-ink principle: most of the ink on a graph should be devoted to the data (the actual content of the graph). Often a graphic can be improved by removing superfluous ink.
 - (a) Following the principle of data-ink, describe a useful change to Marey's train schedule.
 - (b) Describe a change to the usual histogram graphic (such as the one you made in lab) which would increase its proportion of data-ink, without sacrificing readability. Sketch an example. (Hint: the 'content' of a histogram is only the heights and widths of the bars.)
3. From the course web page, download the spreadsheet `hw1.xls`. Load this data into R and make a histogram of the column `PCT40.64`, using 40 bins. This data is the percentage of people aged 40–64 in each census tract in Pennsylvania, according to the 1990 Census. Make sure the axes in the plot are labeled according to this description.
 - (a) Are there exceptional points which lie outside the main distribution?
 - (b) Does the distribution have multiple modes? If so, identify them.
 - (c) Putting together (a) and (b), summarize what this graph says about the geographic distribution of people aged 40–64 in Pennsylvania, in everyday terms. For example, are they uniformly distributed across the state?
4. Find an example in your own life where you could collect data and make use of a histogram to learn something. Try to make it personal enough that no one else in the class would think of it, and make sure that enough data could be collected for the graph to be meaningful.
 - (a) If you found an exceptional value or multiple modes, what would you learn?
 - (b) How could the common mistakes in making histograms affect what you learn?