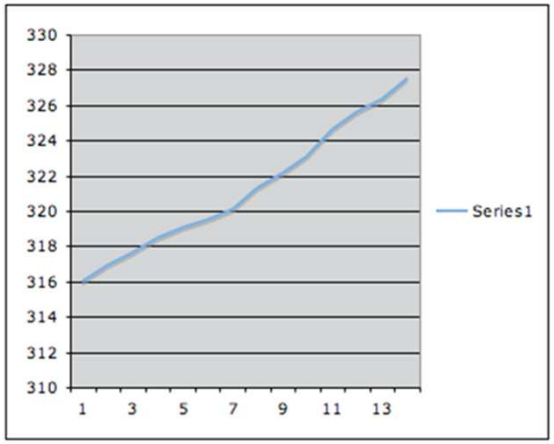
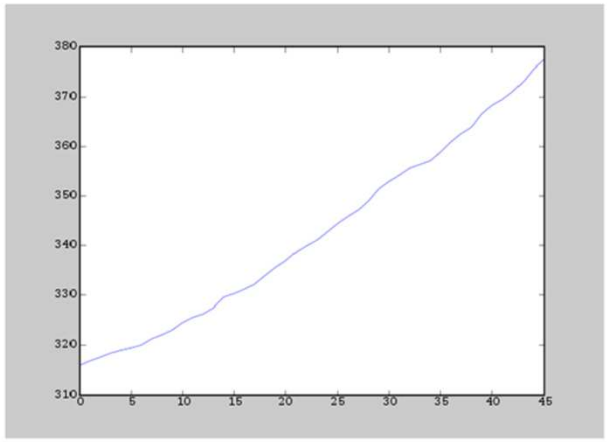
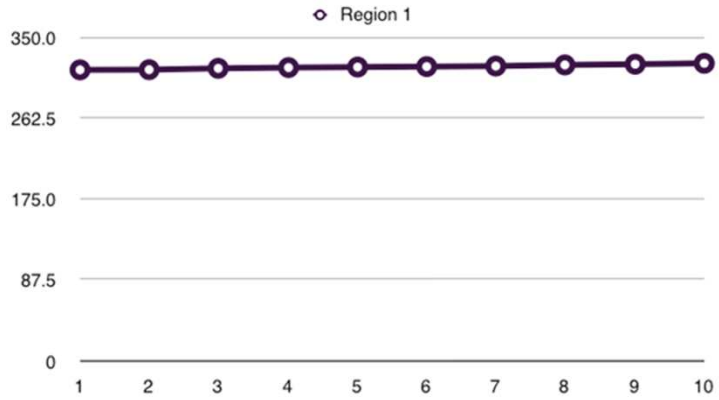
Some Principles for Plots

Visualizing Data [Cleveland 93] and *Elements of Graphing Data* [Cleveland 94] by William S. Cleveland

The information provided here should be considered as guidelines



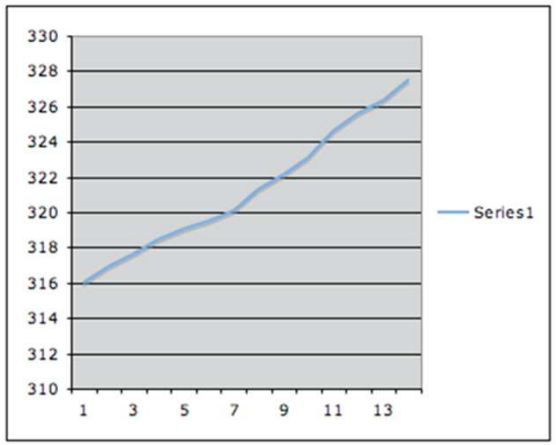


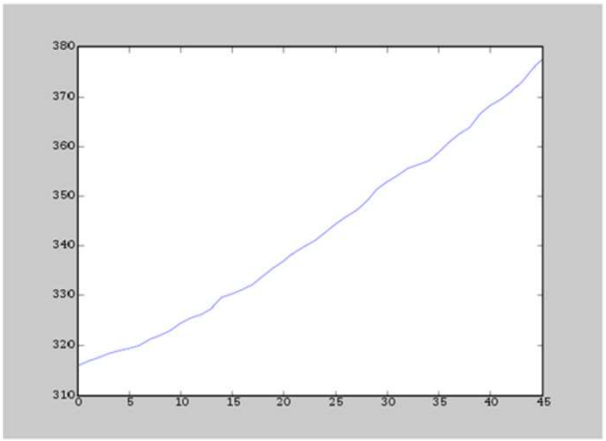


- Why are they all different?
- What is good/bad about each?

Improving the Vision

- Principle 1: Reduced clutter, Make data stand out
  - The main focus of a plot should be on the data itself, any superflous elements of the plot that might obscure or distract the observer from the data needs to be removed.

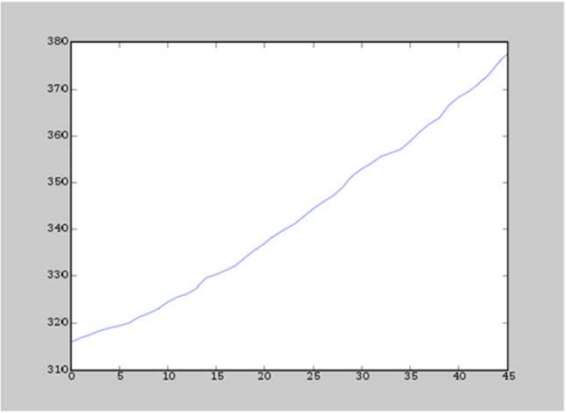


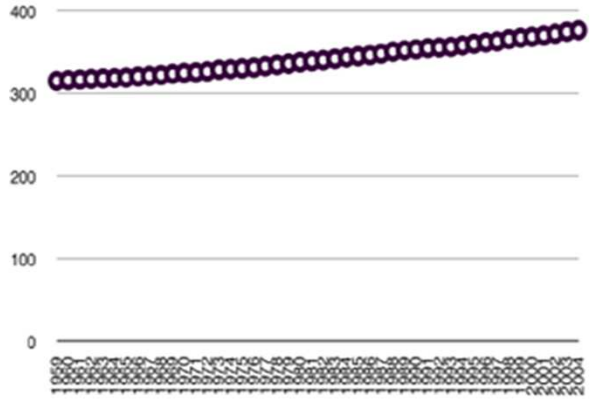


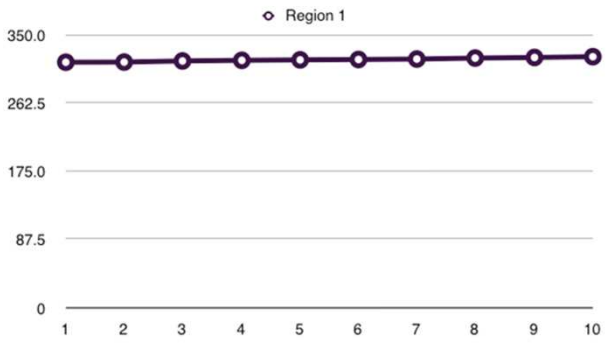
Which one is better?

Improving the Vision

- Principle 2: Use visually prominent graphical elements to show the data.
  - Connecting lines should never obscure points and points should not obscure each other.
  - If multiple samples overlap, a representation should be chosen for the elements that emphasizes the overlap.
  - If multiple data sets are represented in the same plot (superposed data), they must be visually separable.
  - If this is not possible due to the data itself, the data can be separated into adjacent plots that share an axis

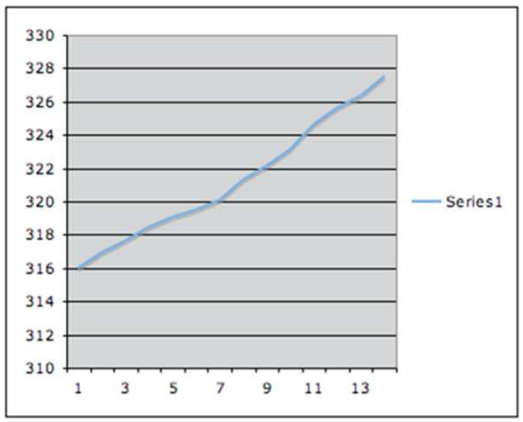


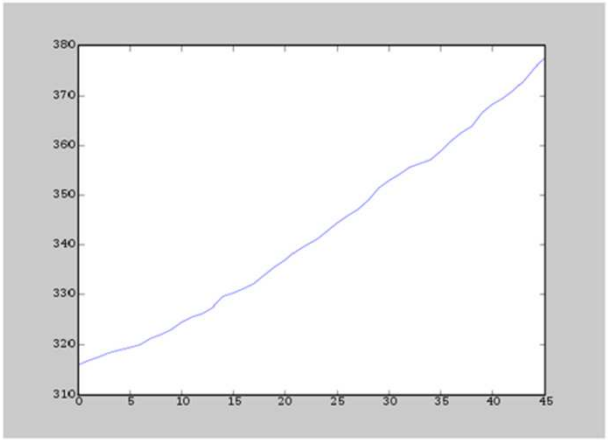


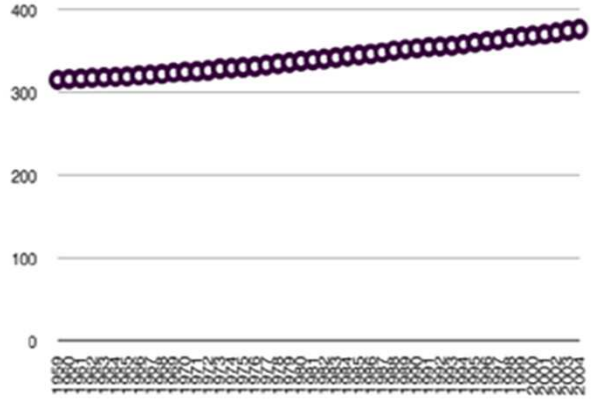


Improving the Vision

- Principle 3: Use proper scale lines and a data rectangle.
  - Two scale lines should be used on each axis (left and right, top and bottom) to frame to data rectangle completely.
  - Add margins for data
  - Tick-marks outs and 3-10 for each axis

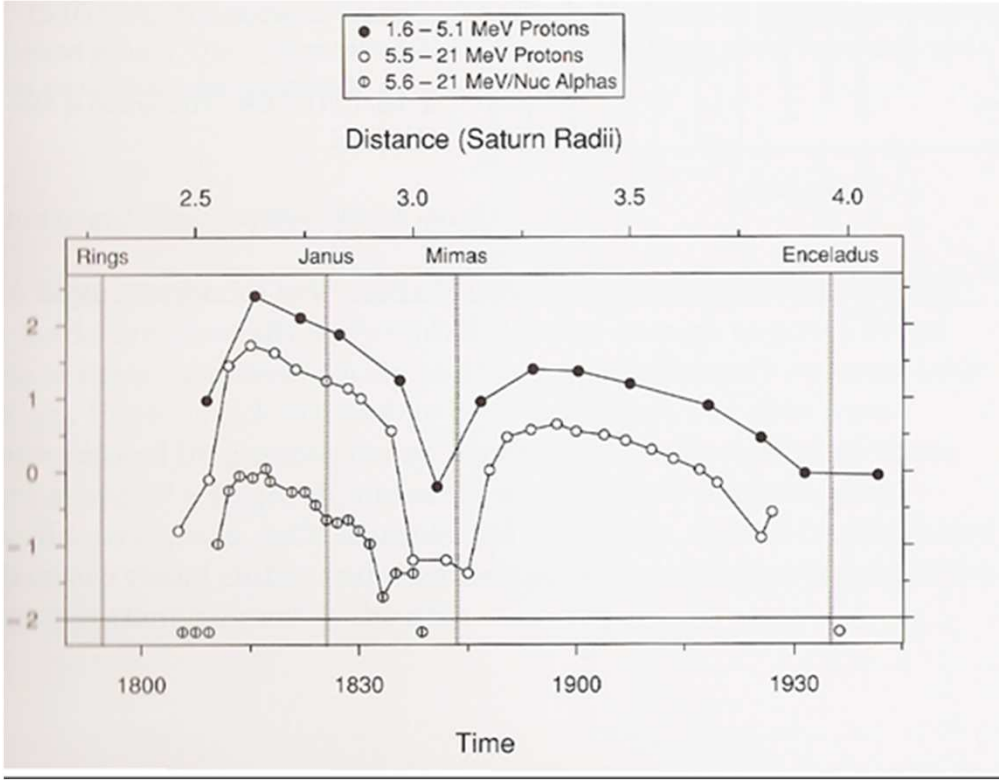






Improving the Vision

- Principle 4: Reference lines, labels, notes, and keys.
  - Only use them when necessary and don't let them obscure data.



Improving the Vision

- Principle 4: Reference lines, labels, notes, and keys.
  - Only use them when necessary and don't let them obscure data.

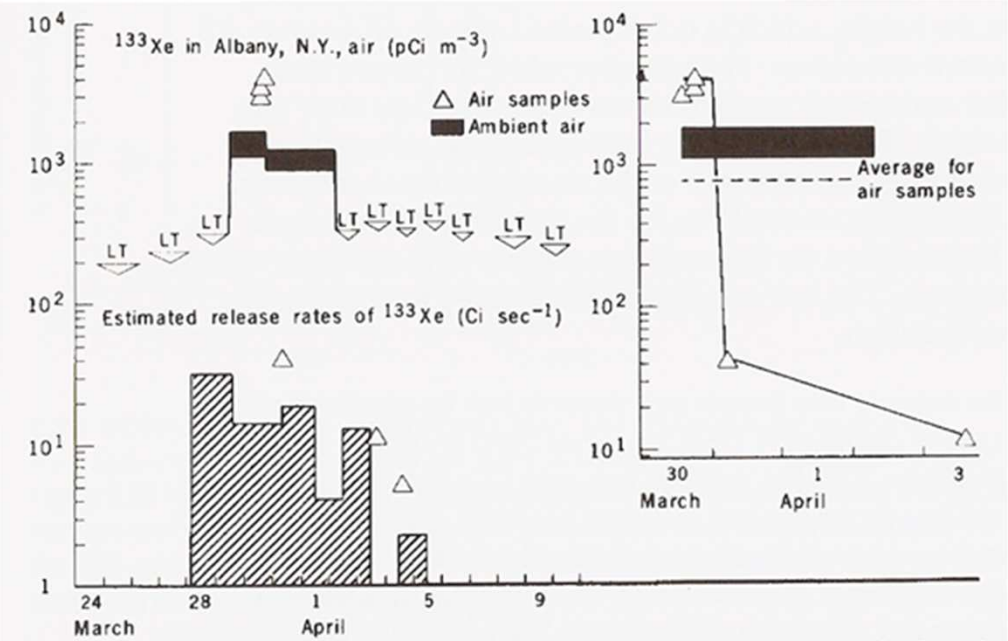
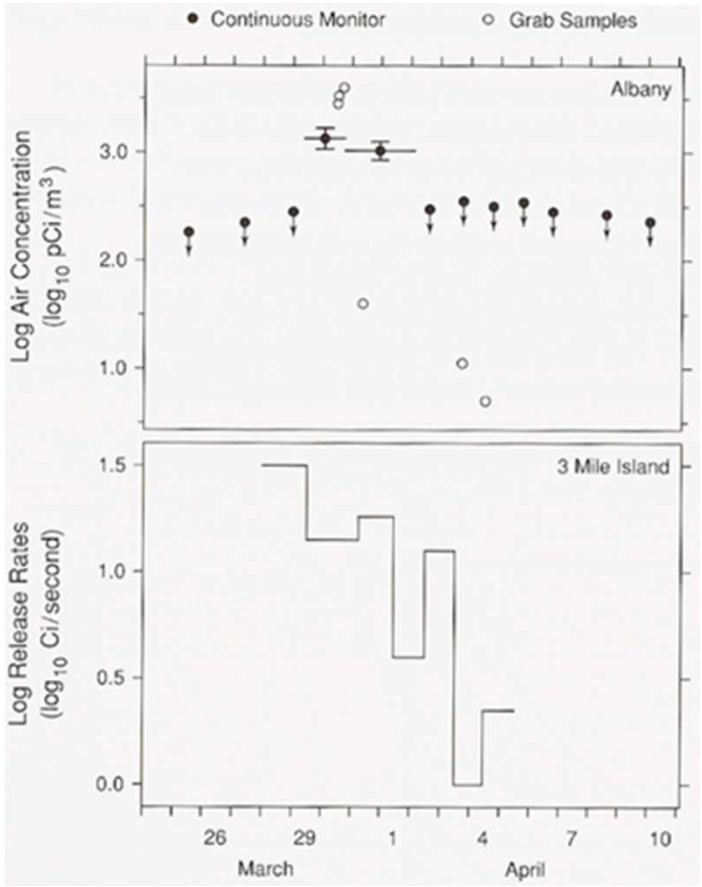
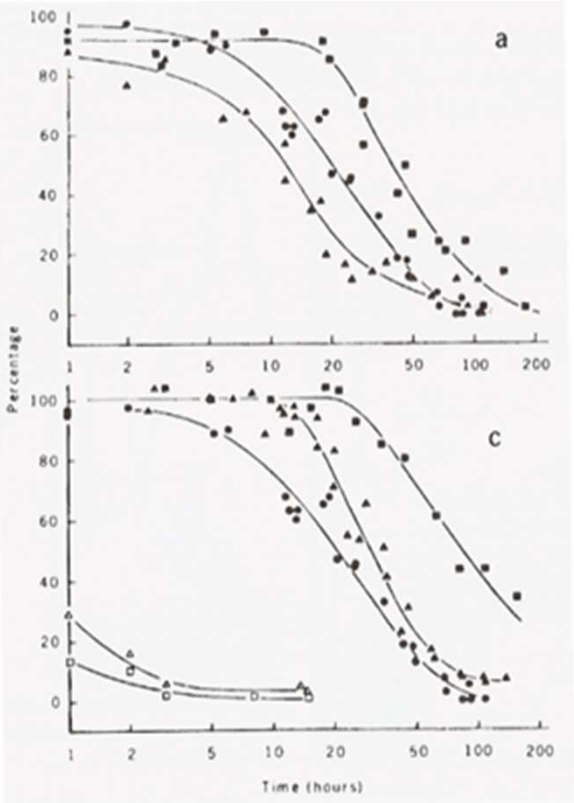


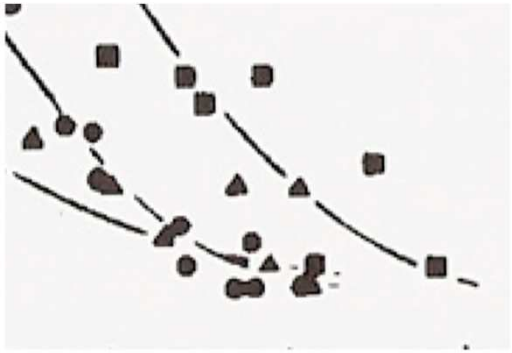
Fig. 1. Xenon-133 activity (picocuries per cubic meter of air) in Albany, New York, for the end of March and early April 1979. The lower trace shows the time-averaged estimates of releases (curies per second) from the Three Mile Island reactor (2). The inset shows detailed values for air samples (gas counting) and concurrent average values for ambient air (Ge diode). Abbreviation: *LT*, less than.



Improving the Vision

- Principle 5: Superposed data set
  - Symbols should be separable and data sets should be easily visually assembled.





Improving the Understanding

- Principle 1: Provide explanations and draw conclusions
  - A graphical representation is often the means in which a hypothesis is confirmed or results are communicated.
  - Describe everything, draw attention to major features, describe conclusions

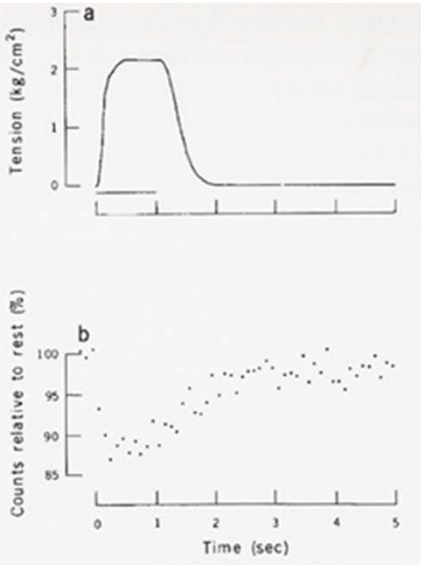
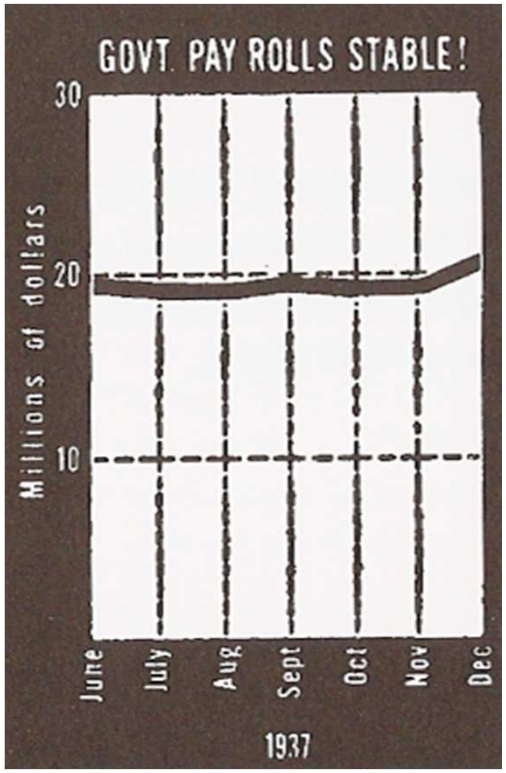


Fig. 2. Tension and the intensity of the 42.9-nm layer line during 1-second tetanus at the sarcomere length of 2.2 μm. (a) Tension record averaged over the 40 tetanic contractions required for obtaining the time course of the layer-line intensity. A sartorius muscle was dissected from *Rana catesbeiana* and tetanized for 1 second at 2-minute intervals. The horizontal line represents the period of stimulation. Tension was recorded with an isometric tension transducer (Shinkoh, type UL). (b) Intensity of the first-order myosin layer line at 42.9 nm. The x-ray source was a rotating-anode generator (Rigaku FR) with a fine focus (1.0 by 0.1 mm) on a copper target. This was operated at 50 kV with a tube current of 70 mA; such a high power was possible with an anode of a large diameter (30 cm) rotating at a high speed (9000 rev/min). A bent-crystal monochromator was used at a source-to-crystal distance of 25 cm with a viewing angle of 6°. The intensity of the myosin layer line was measured with a scintillation counter combined with a mask; the mask had apertures at the positions of the off-meridional parts of the first-order layer line. The meridional reflection at 14.3 nm is known to be slightly displaced during contraction, suggesting a minute change in the myosin periodicity (*1, 3*). It is, therefore, possible that the 42.9-nm layer line is also slightly displaced. However, the possible displacement (14 μm at the position of the mask) would be insignificant compared with the width of each aperture (0.8 mm). The intensity measured at the resting state was 1400 count/sec. The intensities during and after tetanus were expressed as percentages of the resting intensity and plotted against time after the first stimulus of each set of stimuli. Each point represents the intensity averaged over a 100-msec period. The first three points represent the measurements made before stimulation.

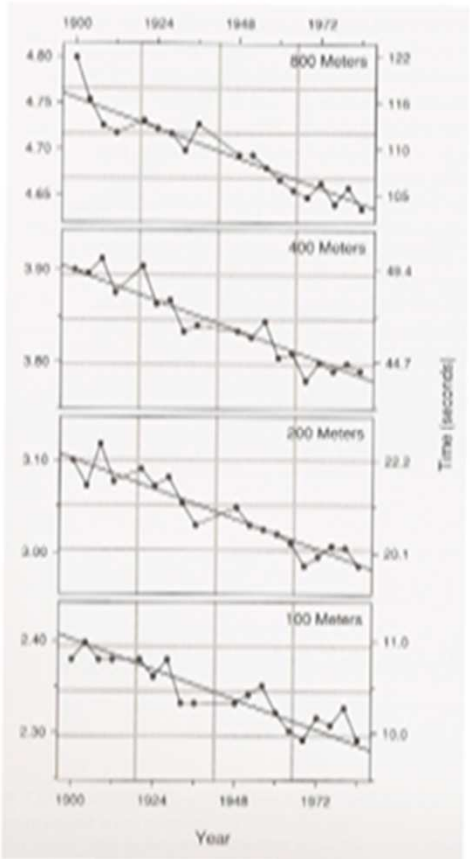
Improving the Understanding

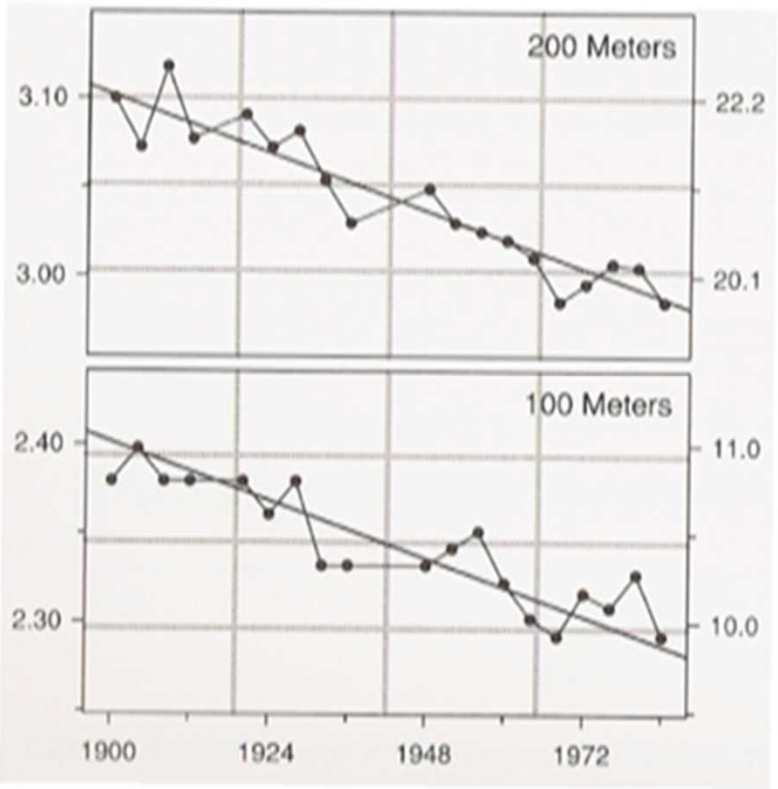
- Principle 2: Use all available space.
  - Fill the data rectangle, only use zero if you need it



Improving the Understanding

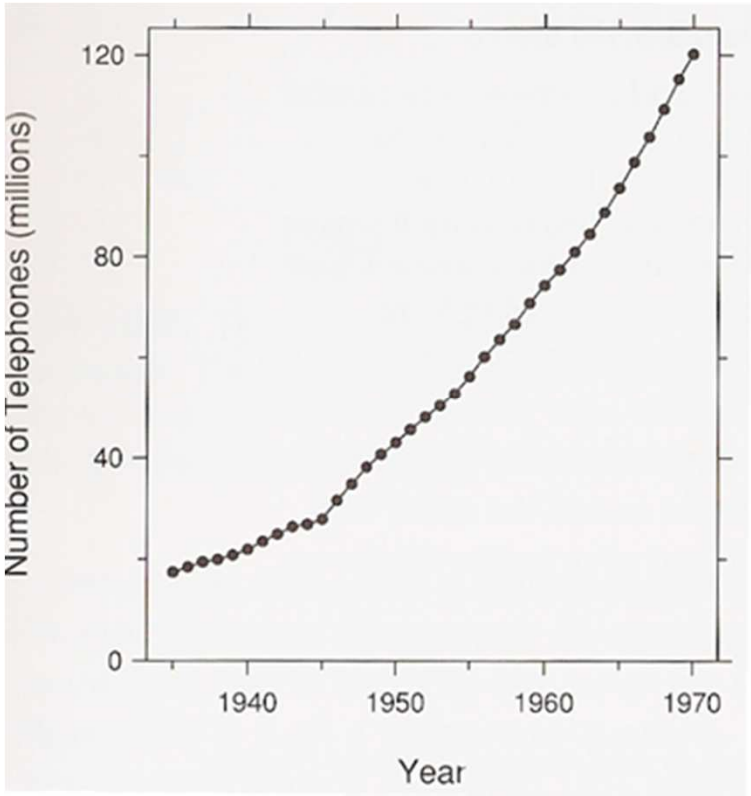
- Principle 3: Align juxtaposed plots
  - Make sure scales match and graphs are aligned

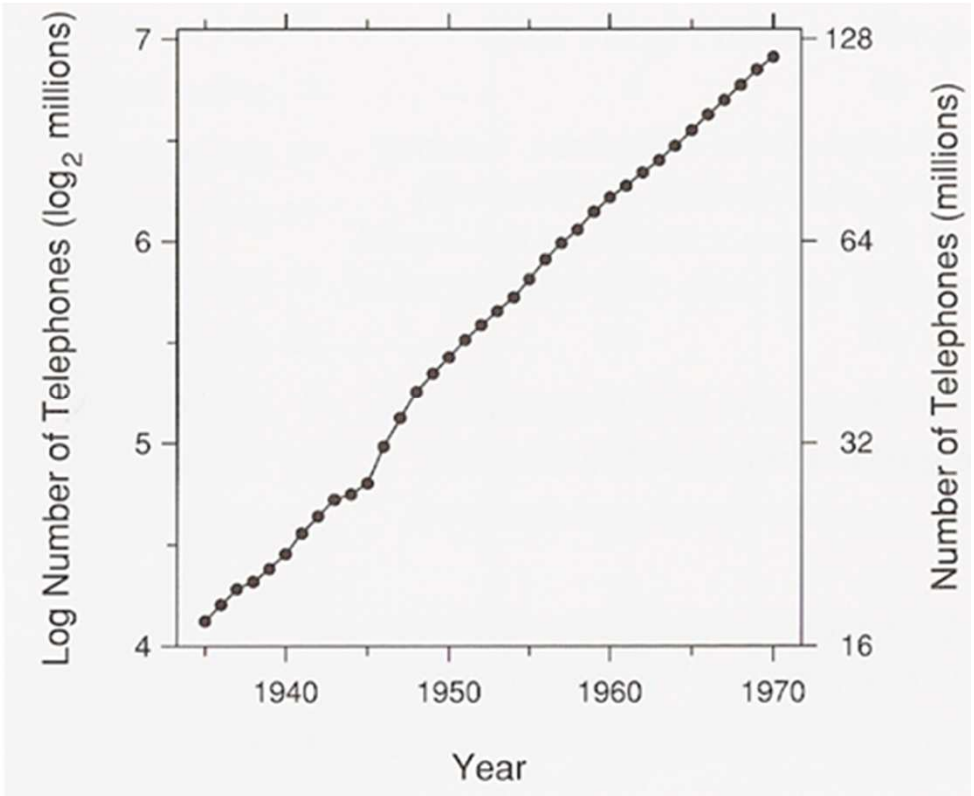




Improving the Understanding

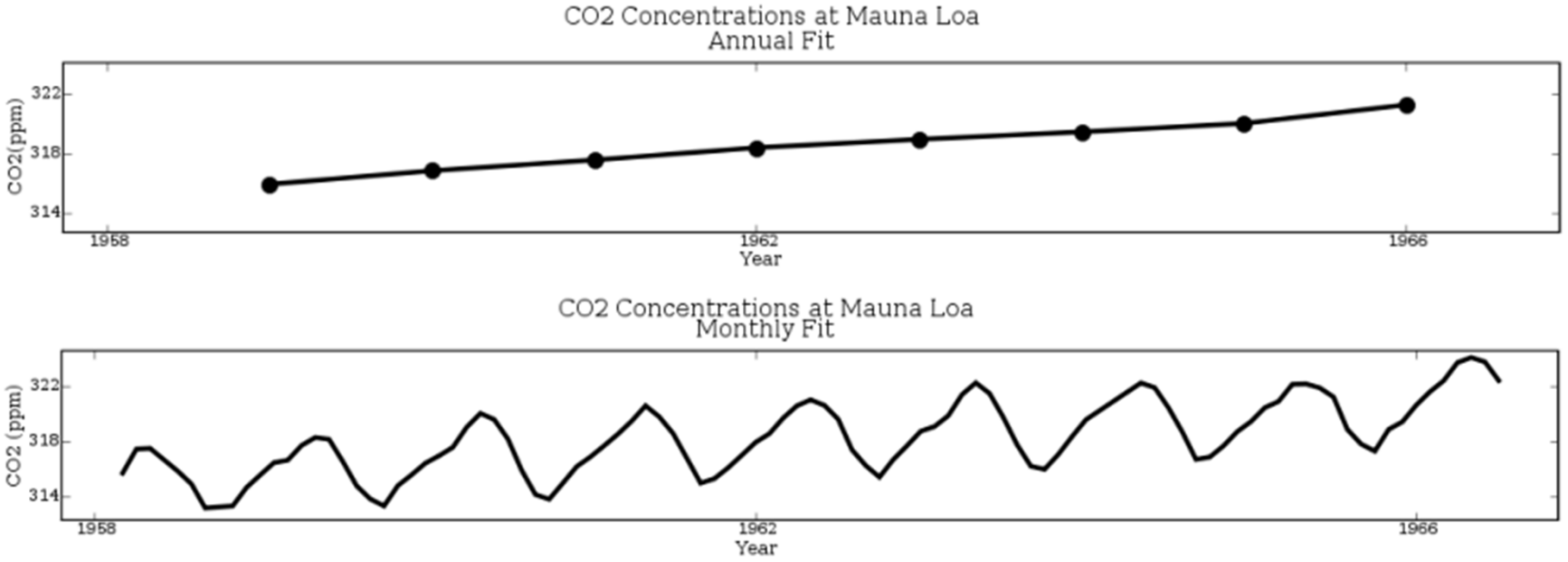
- Principle 4: Use log scales when appropriate
  - Used to show percentage change, multiplicative factors and skewness





Improving the Understanding

- Principle 5: Bank to 45^o
  - Optimize the aspect ratio of the plot



Summary of Principles

- Improve vision
  1. Reduced clutter, Make data stand out
  2. Use visually prominent graphical elements
  3. Use proper scale lines and a data rectangle
  4. Reference lines, labels, notes, and keys
  5. Superposed data set

- Improve understanding
  1. Provide explanations and draw conclusions
  2. Use all available space
  3. Align juxtaposed plots
  4. Use log scales when appropriate
  5. Bank to 45^o