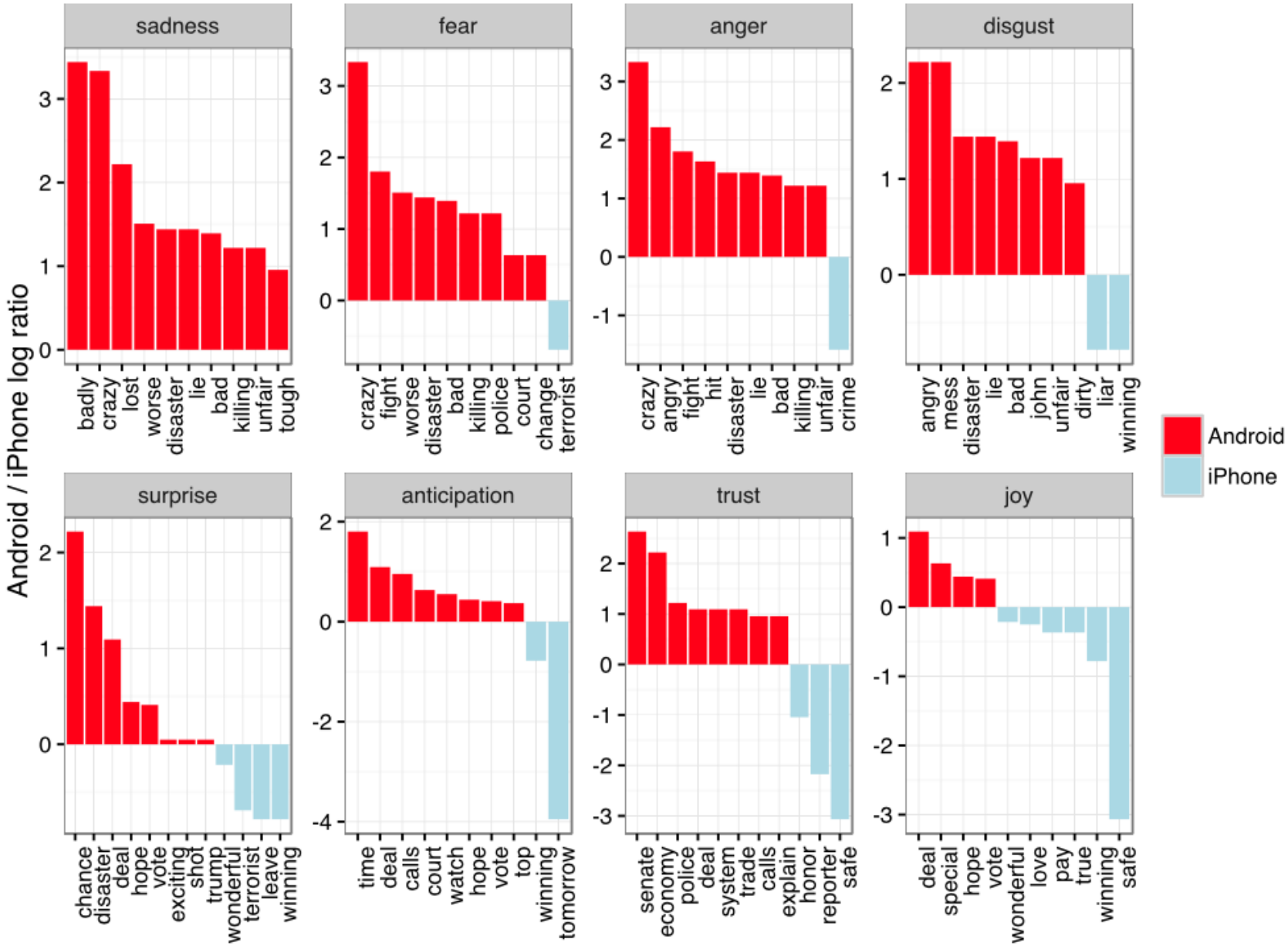
Introduction to Data Visualization

Author: Nicholas G Reich

This material is part of the **statsTeachR** *project*

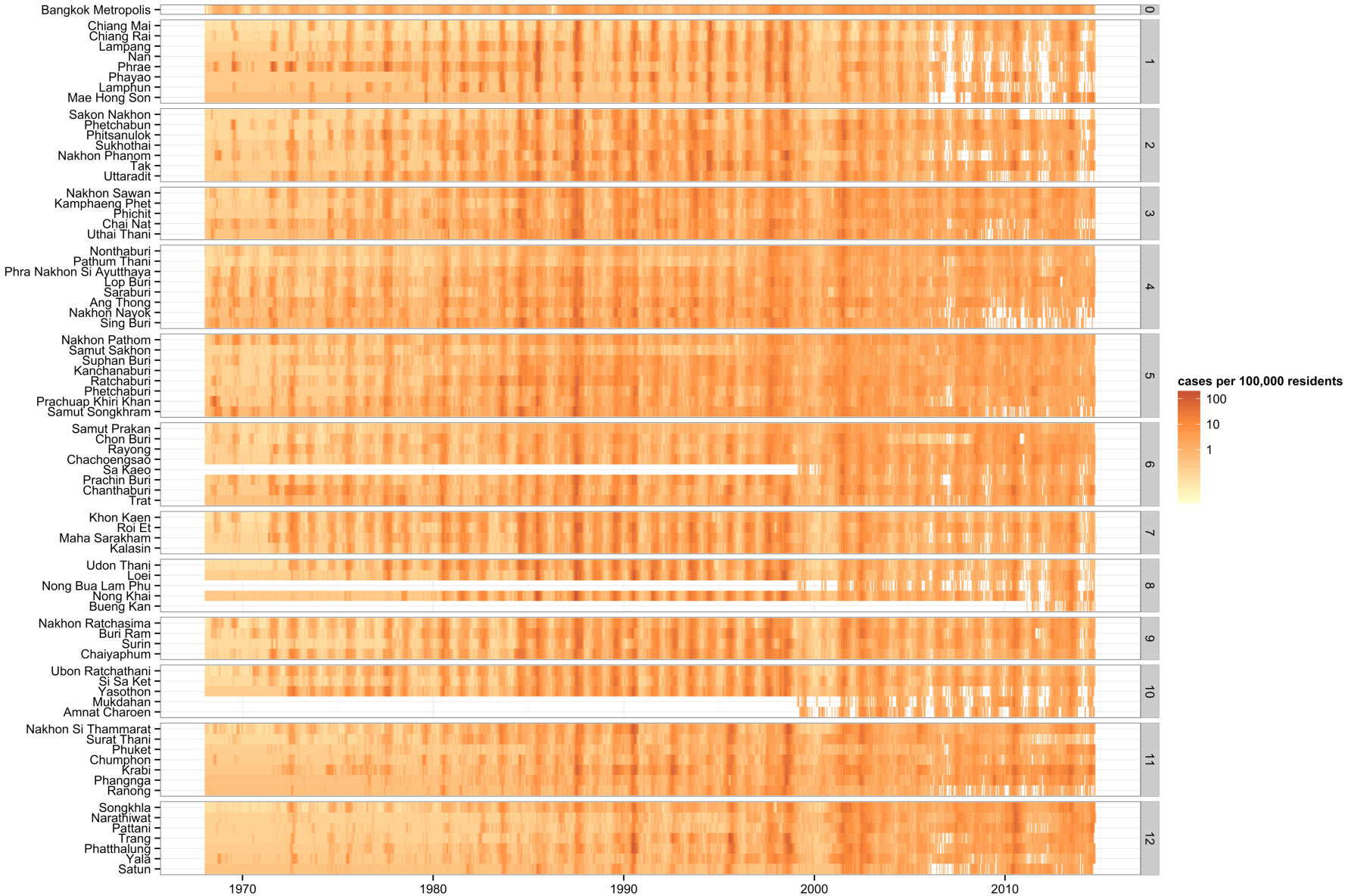
For each of the following graphics, work in pairs to
(1) identify the variables displayed, and
(2) establish two features of the graphic that you like and two that you don't.

Trump tweets[1]



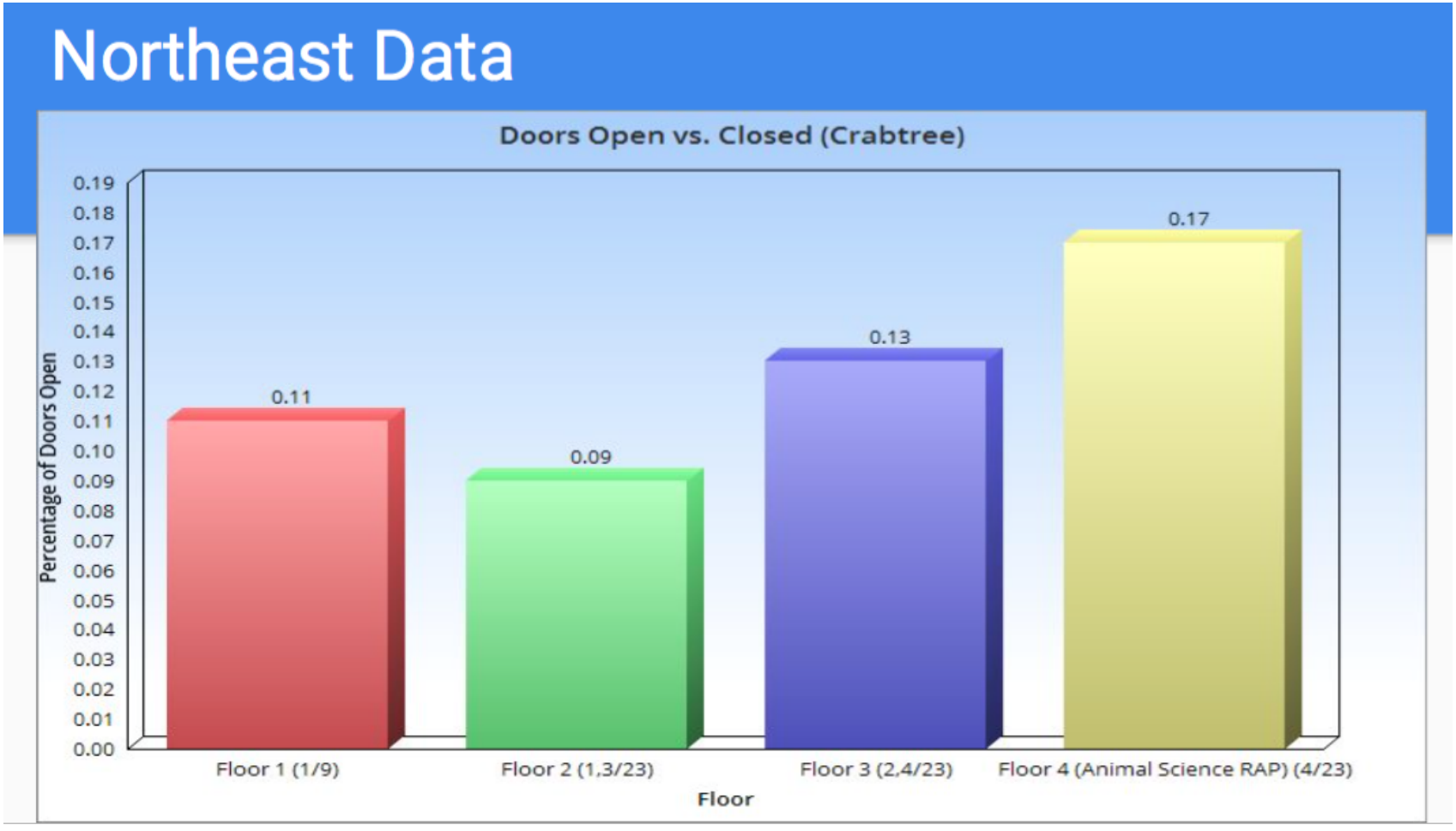
[1] http://varianceexplained.org/r/trump-tweets/

Dengue cases in Thailand[2]



[2] adapted from Reich et al, 2016.

RAP analysis



RAP analysis[3]

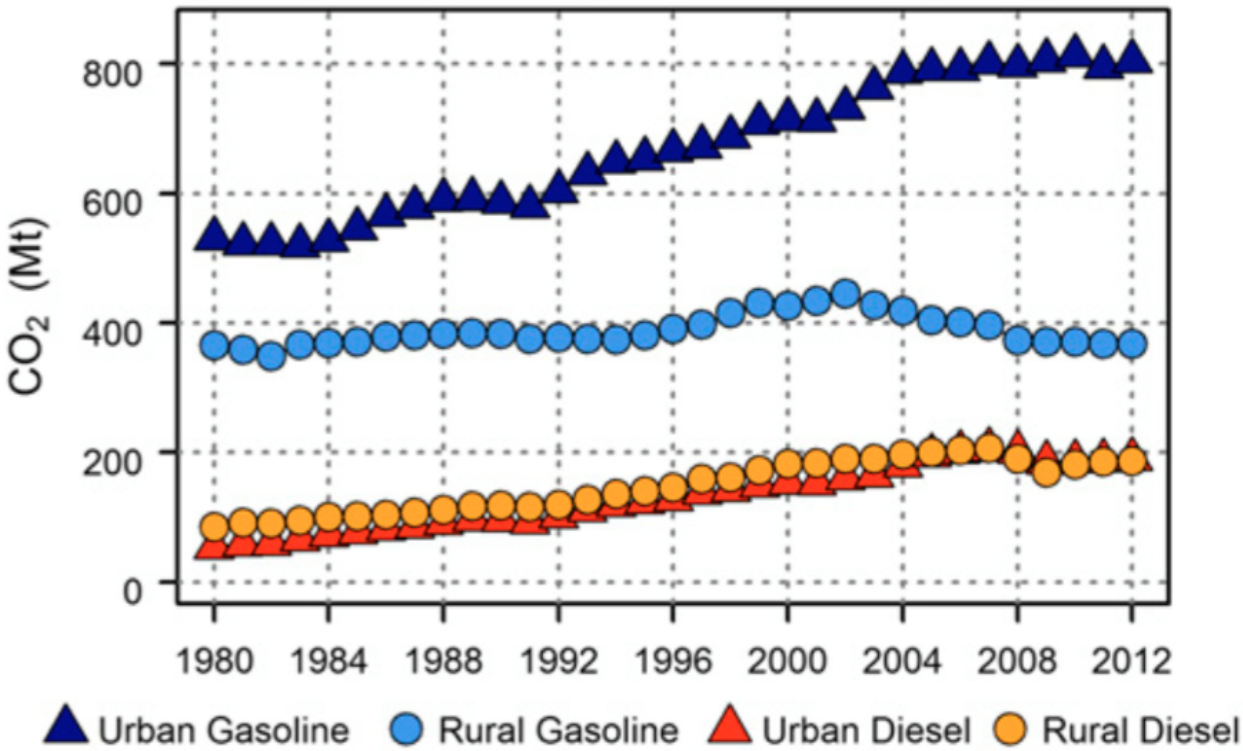


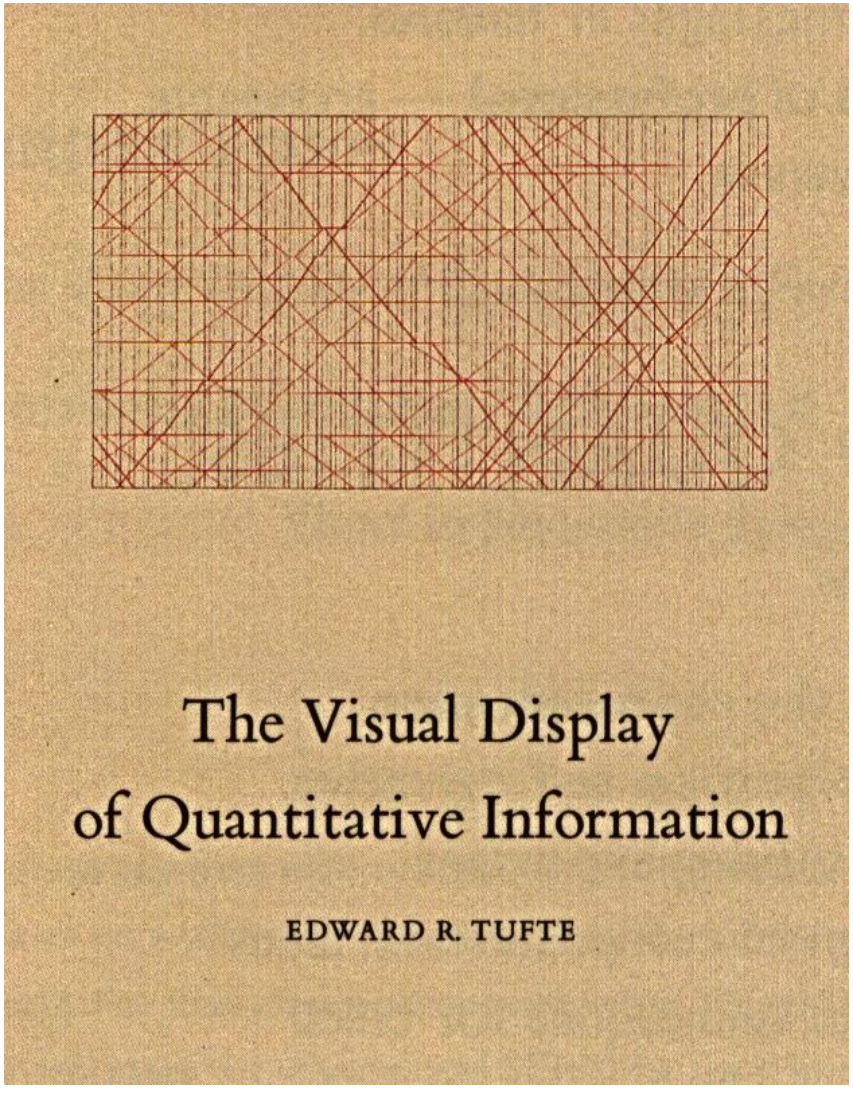
Fig. 2. Time series of US on-road CO_2 emissions. Urban roads accounted for 80% of total emissions growth since 1980. Rural road emissions have been declining since 2002.

[3] from "Cities, traffic, and CO2: A multidecadal assessment of trends, drivers, and scaling relationships", Gately et al, PNAS, 2015.

Why do we visualize data?

Visualization excellence

- consists of complex ideas communicated with clarity, precision, and efficiency.
- is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.
- is nearly always multivariate.
- requires telling the truth about the data.



Key principles of data graphics

- "**Show** the data"
- "encourage the eye to **compare** different pieces of data"
- **Simplify** by maximizing the "data-ink ratio."
- Leverage color, shapes, facets to highlight multivariate data.
- Annotate your figures with context.