Use of R for Visual Display of Public Health Data

Michael Samuel, DrPH
Senior Epidemiologist/Data Scientist
California Department of Public Health

Data

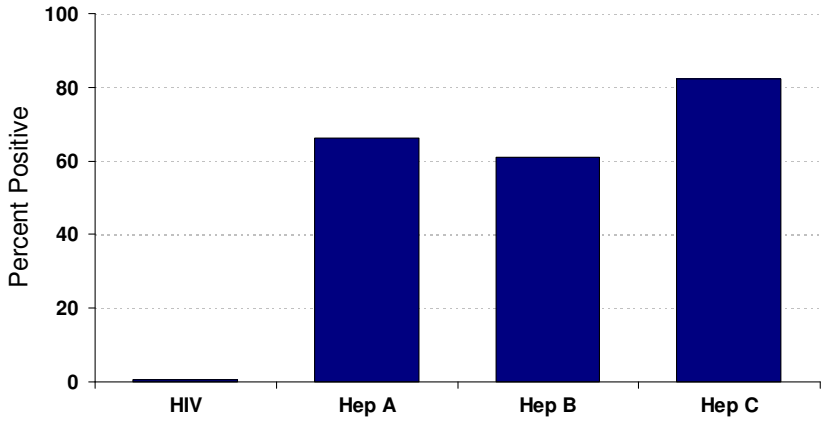
Action

- Program
  - New program
  - Revised program priorities
- New guidelines
- New policy
- New hypothesis (may lead to new action)
- More (or less) money!

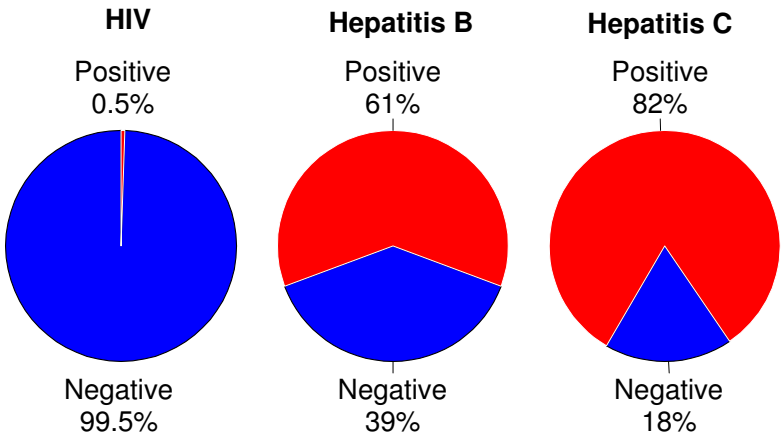
HIV and Hepatitis among Injection Drug Users New Mexico, 1997

	N tested	% Positive	95% C.I.
HIV	1002	0.5	.16-1.6
Hep A	696	66.1	62.4-69.6
Hep B	950	61.1	57.9-64.2
Hep C	945	82.2	79.6-84.6

HIV and Hepatitis Among Injection Drug Users New Mexico, 1997

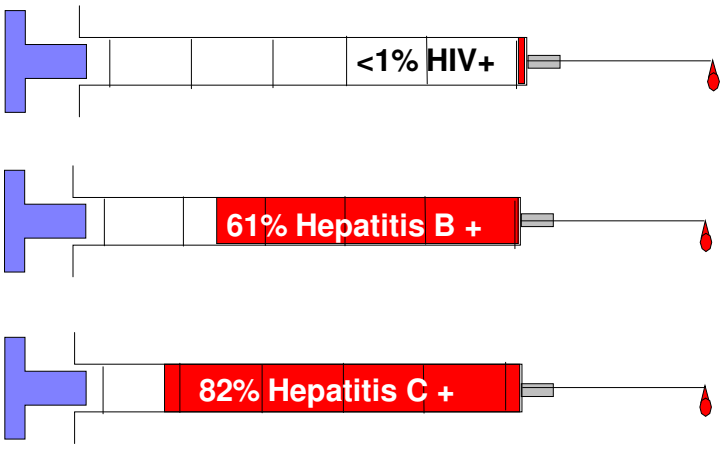


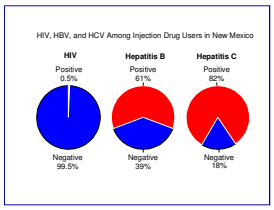
HIV, HBV, and HCV Among Injection Drug Users in New Mexico



HIV, Hepatitis B, and Hepatitis C

New Mexico Street Based Injection Drug User Studies





Data

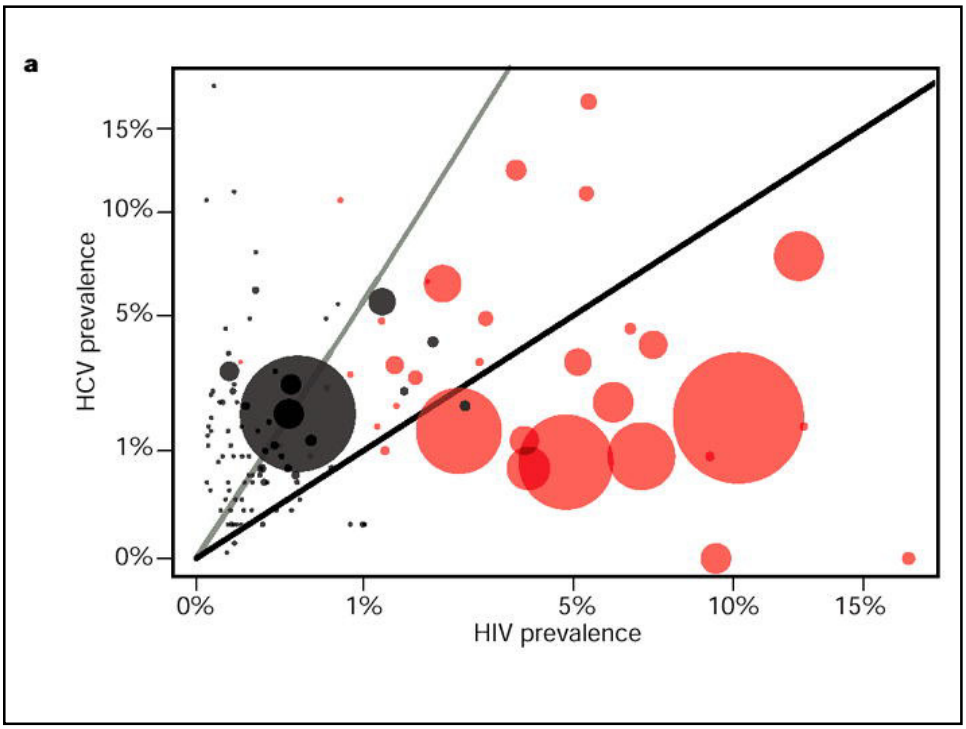
Action

New Mexico is leading the way when it comes to its drug addicts.

Guidelines for Effective Visual Display

- Complexity is good, and…
- Keep it simple, stupid
- Know your audience
- Data integrity
- Clear labels and annotations
- Use appropriate scale(s)
- Use appropriate type of chart
- Pay attention to details
- Less is more
- Avoid extraneous “Chart Junk”


NEW SECOND EDITION JUST PUBLISHED
SECOND EDITION
The Visual Display of Quantitative Information
EDWARD R. TUFTE
Edward R. Tufte
Envisioning Information
EDWARD R. TUFTE
VISUAL EXPLANATIONS
IMAGES AND QUANTITIES, EVIDENCE AND NARRATIVE
Edward R. Tufte
The Cognitive Style of PowerPoint

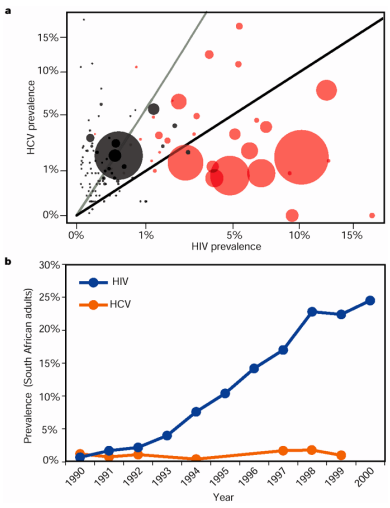

a
15%
10%
5%
1%
0%
HCV prevalence
0%
1%
5%
10%
15%
HIV prevalence

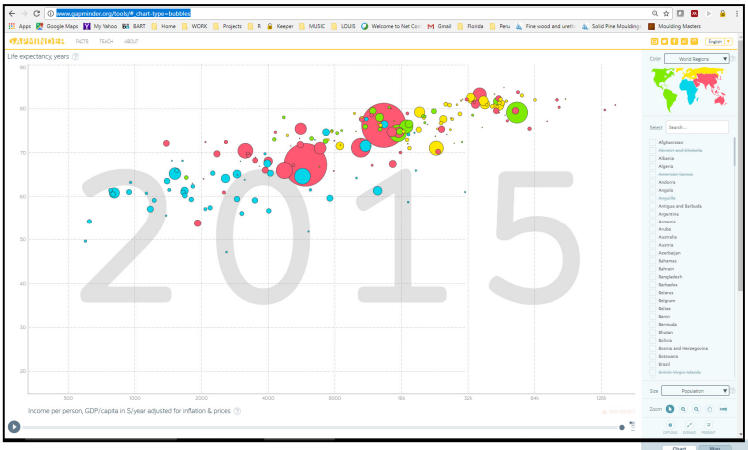
Nature **422**, 679 (17 April 2003)

Epidemiology: Sexual transmission of HIV in Africa

Polly R. Walker, Michael Worobey, Andrew Rambaut, Edward C. Holmes and Oliver G. Pybus

Figure 1 Comparison of epidemic histories of human immunodeficiency virus (HIV) and hepatitis C virus (HCV) in different countries, and in South Africa during the 1990s. **a,** HCV and HIV prevalence in the general population (including children) of every country for which data are available. The arcsine square-root transformation for proportions was used because the original distribution of points was strongly L-shaped. Sub-Saharan African countries are represented by red circles; other countries are represented by black circles; the radius of each circle is proportional to the number of HIV infected individuals in that country. Countries below the black line have a higher prevalence of HIV than HCV; the opposite is true for countries above the black line. The grey line represents an HCV:HIV prevalence ratio of 6, which indicates the relative parenteral transmissibility of the two pathogens 3. All prevalence figures apart from five HCV values were obtained from World Health Organization sources. **b,** HIV and HCV prevalence in South African adults from 1990 to 2000 (estimates for HIV from ref. 6; estimates for HCV from sources listed in supplementary information).





http://www.gapminder.org/tools/

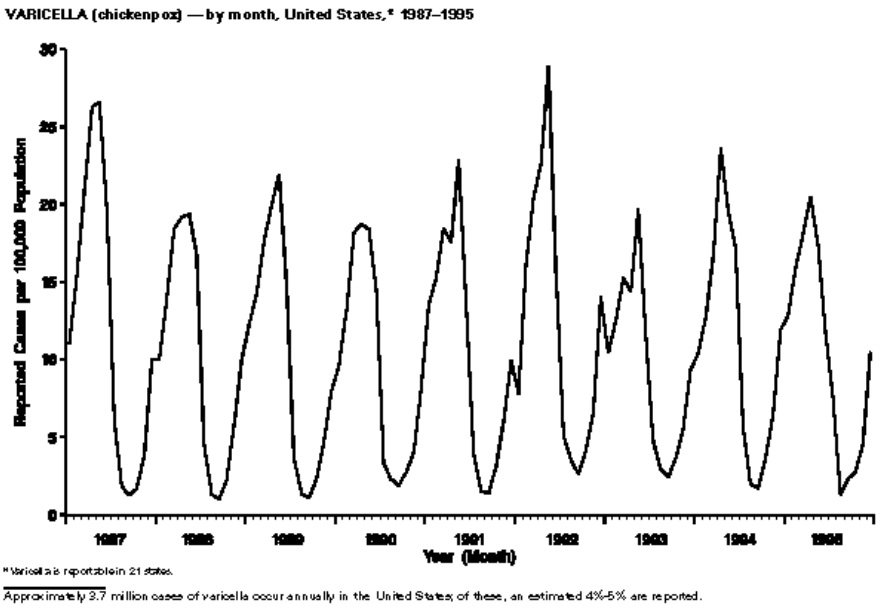
https://www.nytimes.com/interactive/2018/03/19/upshot/race-class-white-and-black-men.html
https://www.nytimes.com/interactive/2018/03/27/upshot/make-your-own-mobility-animation.html

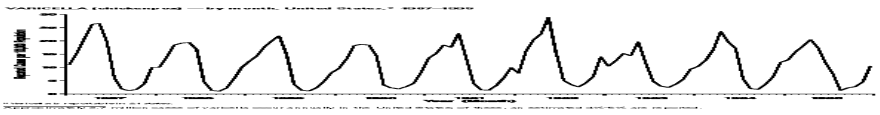
Display Types

- Tables
- Charts
  - Line (simple, multiple)
  - Bar (simple, clustered, stacked, 100%; vertical, horizontal)
  - Pie
- Plots
  - Histograms
  - Box
  - Scatter
  - Regression-related
- Maps
- And many others including combinations the above

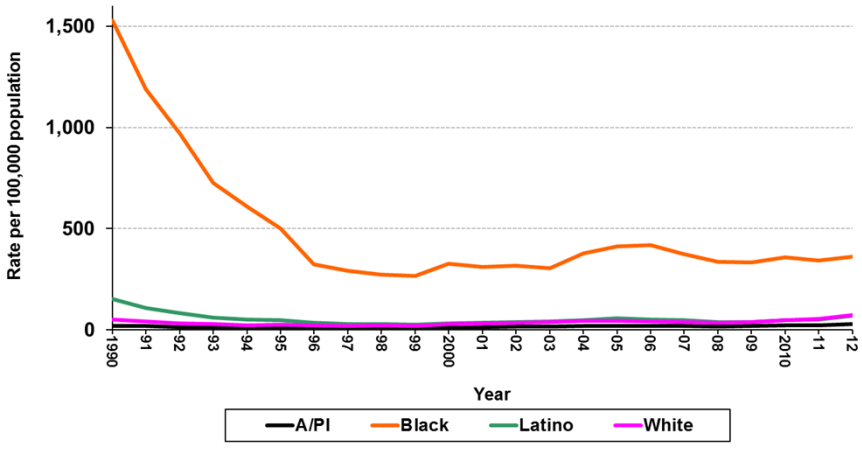
"Nut and Bolts"

- Scale and Proportion
- Labels and Legends
- Grid Lines
- Color
- Animation/"PowerPoint"
- Font
- 3D
- Production/Reproduction
- Chart Junk
- Software



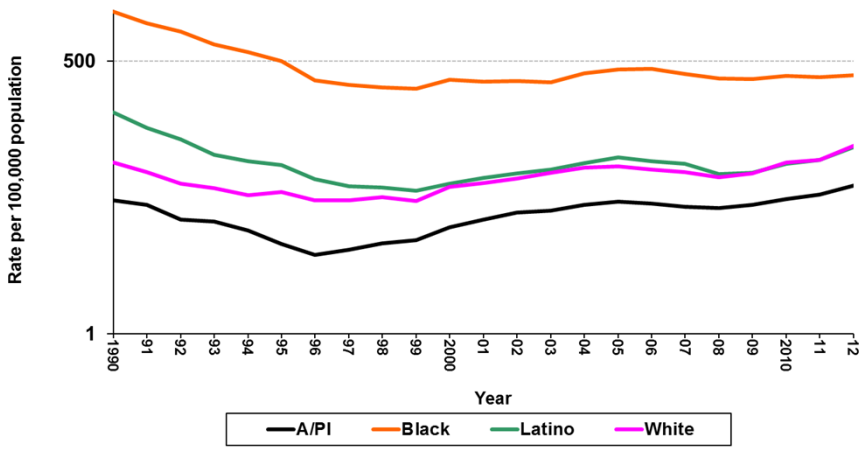


Gonorrhea, Rates for Males by Race/Ethnicity
California, 1990–2012



Note: NA/AN = Native American/Alaskan Native, A/PI = Asian/Pacific Islander.
Race/ethnicity "Not Specified" ranged from 21.1% to 36.1% of cases for males in any given year.

Gonorrhea, Rates for Males by Race/Ethnicity
California, 1990–2012



Note: NA/AN = Native American/Alaskan Native, A/PI = Asian/Pacific Islander.
Race/ethnicity "Not Specified" ranged from 21.1% to 36.1% of cases for males in any given year.

For More Information:

Part 1 "General Concepts"
- Part 1a: http://youtu.be/1c41eMOjt_U
- Part 1b: http://youtu.be/XlKA2hgg-rY

Part 2 "Nuts and Bolts"
- Part 2a: http://youtu.be/pUDcGlulfW8
- Part 2b: http://youtu.be/YCRyVPpz-yk

Why Use R for Visual Display

- Data, calculations/statistics, text and any other information can be placed with code anywhere on any type of visual display
- Mind-blowing array of plotting functionality from core program and contributed packages
- Flexible system for getting data "in"
- Powerful system for organizing, processing, formatting and analyzing data
- Flexible system for getting visual displays out