

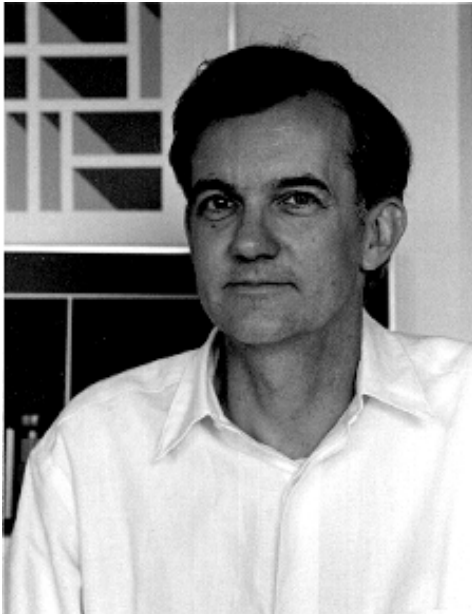
Data Visualization in Biomedical Research



Meridith Blevins, MS
Department of Biostatistics
Vanderbilt Institute for Global Health



Experts in data visualization



Edward Tufte
*The Visual Display of
Quantitative
Information.* 1985.



William S. Cleveland
Visualizing data. 1993.



A four step recipe for creating a graphic



Step 1: Ingredients

Identify Key Variables

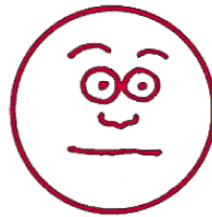
- **There are different types of graphs to support the different scales of measurement**
 - **Nominal Scale**
 - **Ordinal Scale**
 - **Interval Scale**



0
No Hurt



1
Hurts
Little Bit



2
Hurts
Little More



3
Hurts
Even More



4
Hurts
Whole Lot



5
Hurts
Worst

Step 2: Palette



Determine Role of Visualization

- **There are different roles for data visualization**
 - **Presentation vs. Exploratory (then stop here)**
 - **Summary/Descriptive**
 - **Distribution**
 - **Comparison**
 - **Association**
- **Answer these questions**
 - **What to Whom, How, and Why?**

Step 3: The Dish



Choice of Graphical Form

- **Based on the Key Variables (or ingredients) and Role of Visualization (or palette)**
 - **2D plot (most common)**
 - **Multiple 2D plots**
 - **Interactive plot**
 - **Multiple plots stringed together like a story line**
- **We will go through the different graphical forms in our R examples**

Step 4: Add Garnish



- **Think about improvements that could be made to improve clarity and focus**
 - Scales
 - Sorting and Ordering
 - Overlays
 - Positioning in text
 - Size, Frames and Aspect Ratio
 - **Color**

“Ink-to-Information”

= data-ink ÷

total ink used to print the graphic

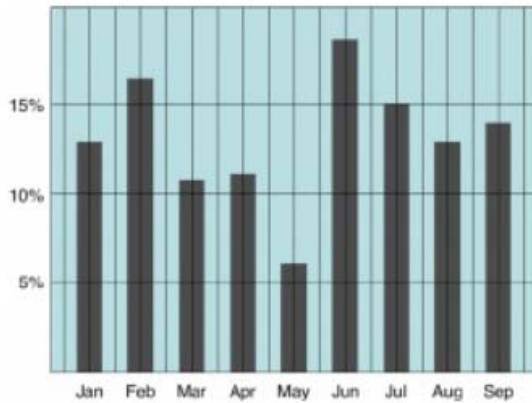
= proportion of graphic's ink devoted to
the non-redundant display of data

= 1 – proportion of graphic that can be
erased

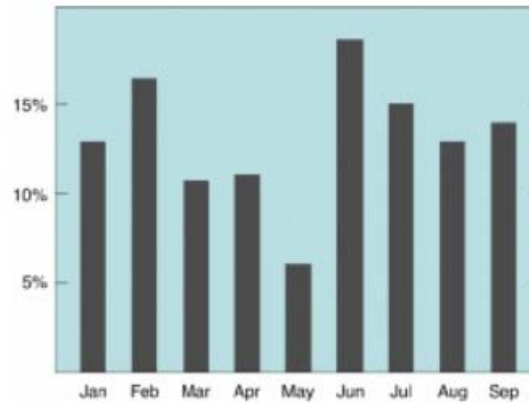
- Also called data-ink ratio

“Ink-to-Information”

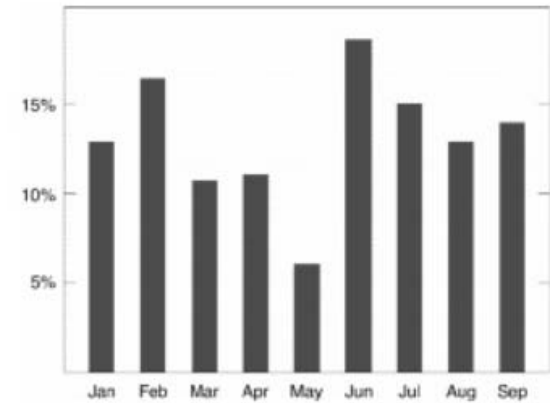
1



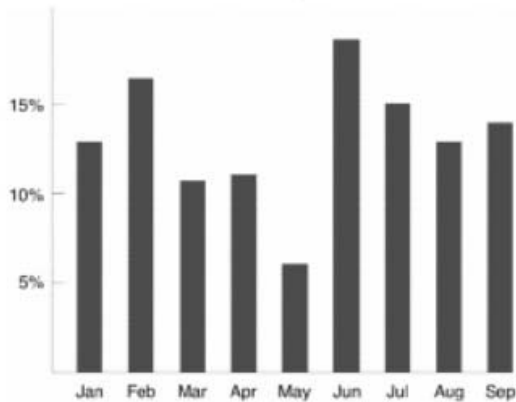
2



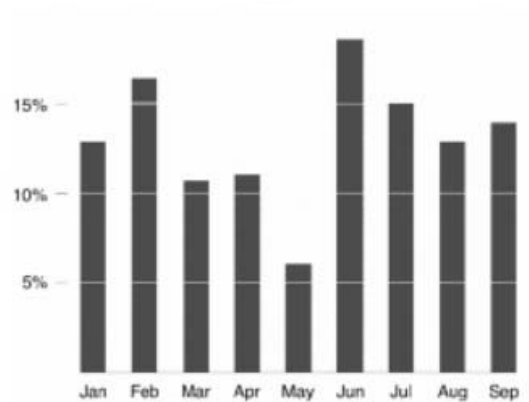
3



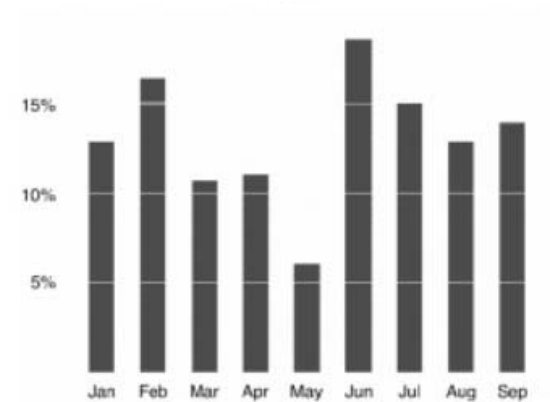
4



5



6



source: The Visual Display of Quantitative Information by Edward R. Tufte

Choosing Colors

colorbrewer2.org

golden ratio

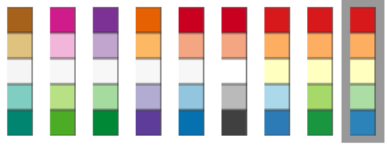
Number of data classes: 8

how to use | updates | downloads | credits

COLORBREWER 2.0
color advice for cartography

Nature of your data:
 sequential diverging qualitative

Pick a color scheme:



colorbrewer2.org

Only show:
 colorblind safe
 print friendly
 photocopy safe

Context:
 roads
 cities
 borders

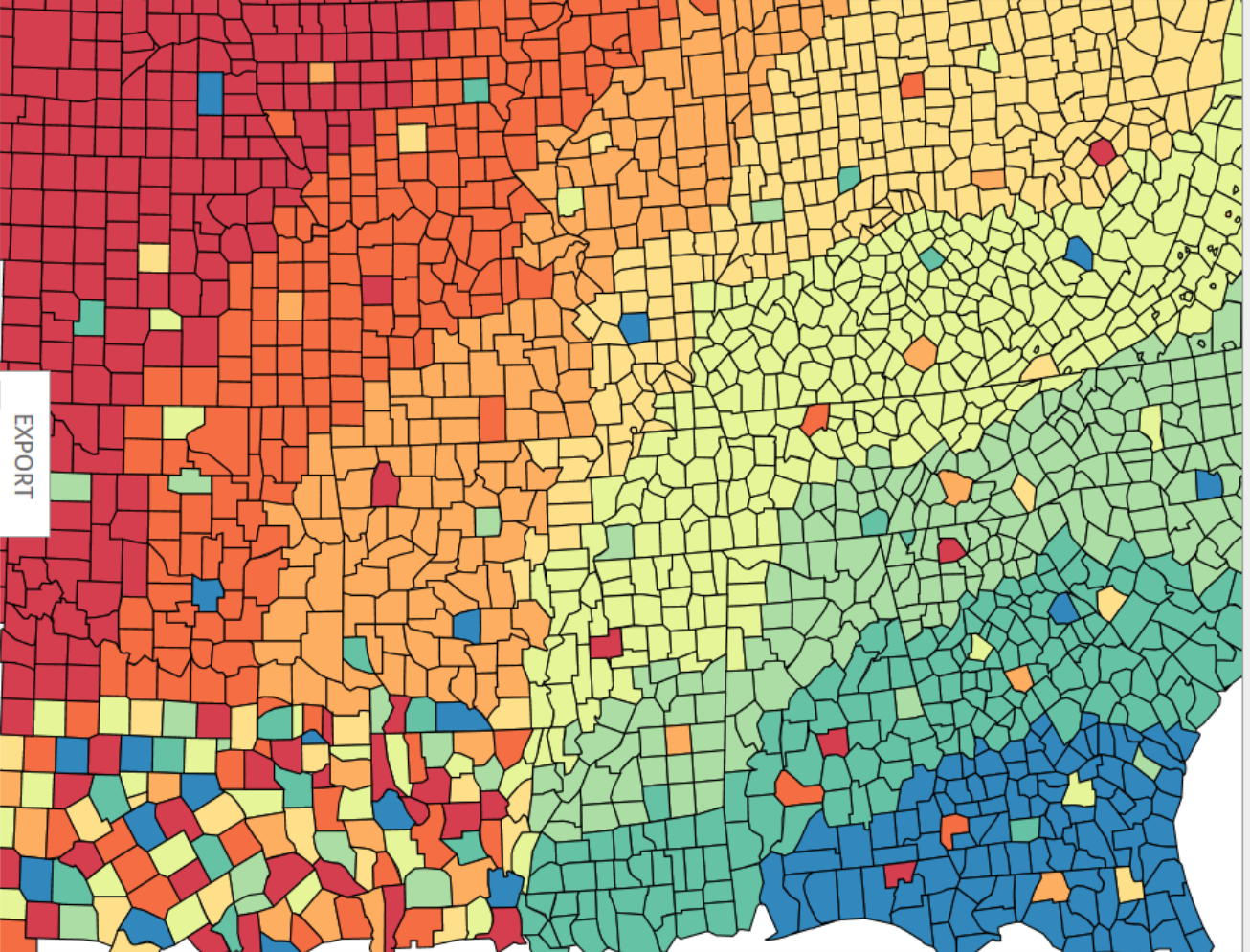
Background:
 solid color
 terrain
color transparency

8-class Spectral

EXPORT

HEX

	#d53e4f
	#f46d43
	#fdae61
	#fee08b
	#e6f598
	#abdda4
	#66c2a5
	#3288bd



Choosing Colors

colorschemedesigner.com/csd-3.5/

Color Scheme Designer is now Paletton! Check out [Paletton.com](#)

Color Scheme Designer → **paletton**

2002-2010 © Petr Stanicek • v3.51 • [Blog & News](#)

mono complement triad tetrad analogic **accented analogic**

Undo Redo Random **Colorblind** Color space Export Help

Scheme ID: 01626w0w0w0w0

CREATE FREE STUNNING WEBSITES GO >>

colorschemedesigner.com

Hue: 45° opposite Angle: 44°

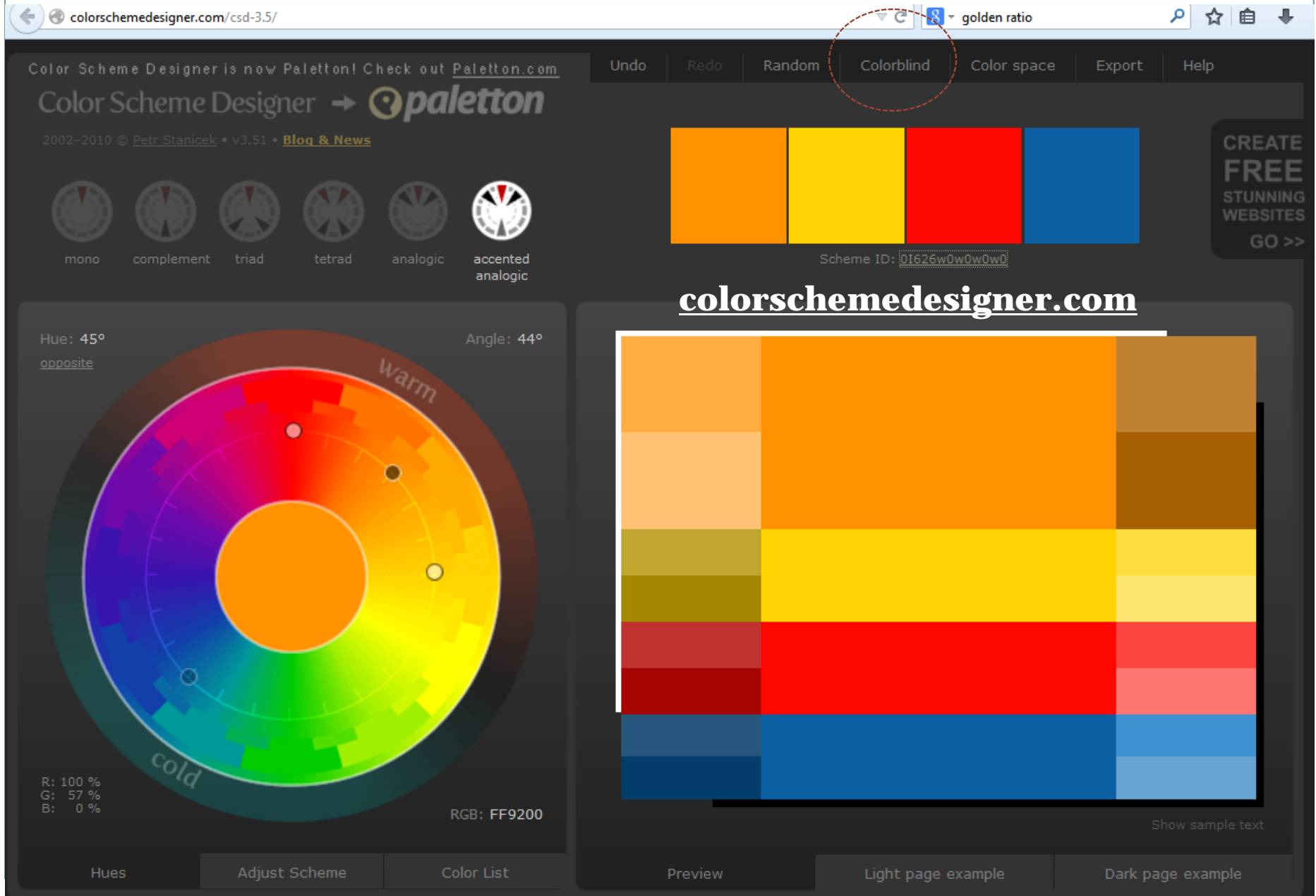
warm cold

R: 100 % G: 57 % B: 0 % RGB: FF9200

Hues Adjust Scheme Color List

Preview Light page example Dark page example

Show sample text

The image is a screenshot of the Color Scheme Designer website. At the top, the browser address bar shows 'colorschemedesigner.com/csd-3.5/' and the page title is 'golden ratio'. The main navigation bar includes 'Undo', 'Redo', 'Random', 'Colorblind' (highlighted with a red dashed circle), 'Color space', 'Export', and 'Help'. Below this, there's a header for 'Color Scheme Designer' with the 'paletton' logo and version information. A row of color palette selection icons is shown, with 'accented analogic' selected. A horizontal bar displays four color swatches: orange, yellow, red, and blue. Below this bar, the 'Scheme ID' is '01626w0w0w0w0'. A large watermark 'colorschemedesigner.com' is overlaid on the right side. On the left, a color wheel is visible with 'Hue: 45° opposite' and 'Angle: 44°'. The wheel is divided into 'warm' and 'cold' regions. Below the wheel, the RGB values are 'R: 100 % G: 57 % B: 0 %' and the hex code is 'RGB: FF9200'. At the bottom, there are buttons for 'Hues', 'Adjust Scheme', and 'Color List'. On the right, there are buttons for 'Preview', 'Light page example', and 'Dark page example', along with a 'Show sample text' link.

Designing for Colorblind Audience

<http://www.color-blindness.com/coblis-color-blindness-simulator/>

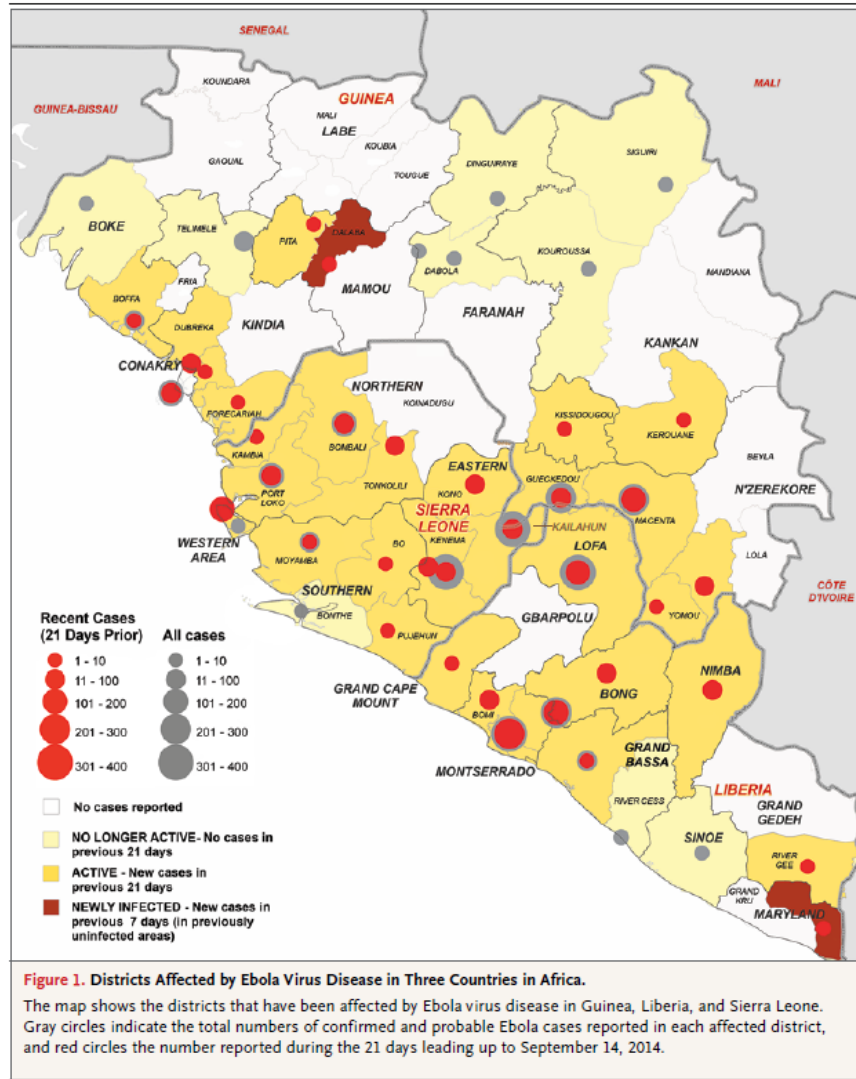
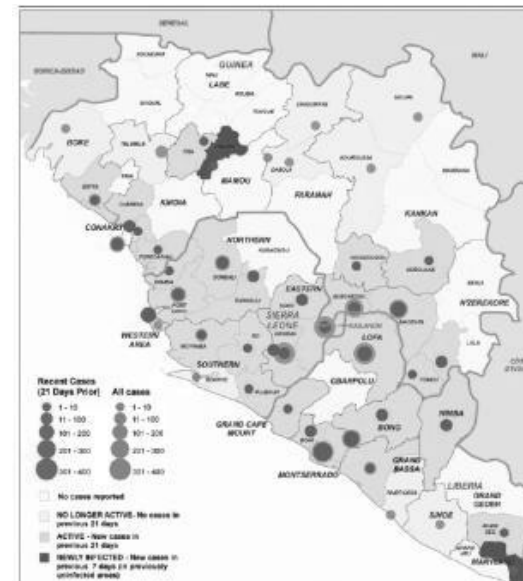


Figure 1. Districts Affected by Ebola Virus Disease in Three Countries in Africa.
 The map shows the districts that have been affected by Ebola virus disease in Guinea, Liberia, and Sierra Leone. Gray circles indicate the total numbers of confirmed and probable Ebola cases reported in each affected district, and red circles the number reported during the 21 days leading up to September 14, 2014.

- ↳ Protanopia (red-blind)
- ↳ Deuteranopia (green-blind)
- ↳ Tritanopia (blue-blind)
- ↳ Protanomaly (red-weak)
- ↳ Deuteranomaly (green-weak)
- ↳ Tritanomaly (blue-weak)
- ↳ Achromatopsia (monochromacy)
- ↳ Blue Cone Monochromacy"



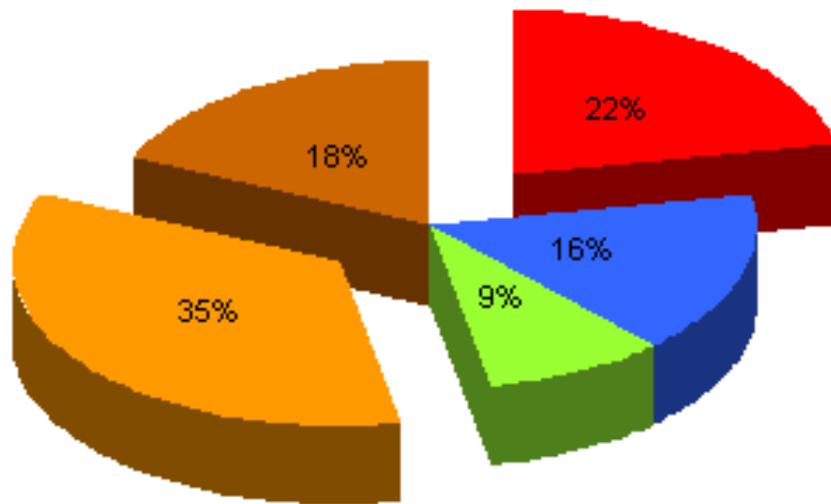
In USA,
 estimated:
 ♂ 7%
 ♀ 0.4%

This map was published on September 23, 2014, at NEJM.org

Bad Practices

Total Sales

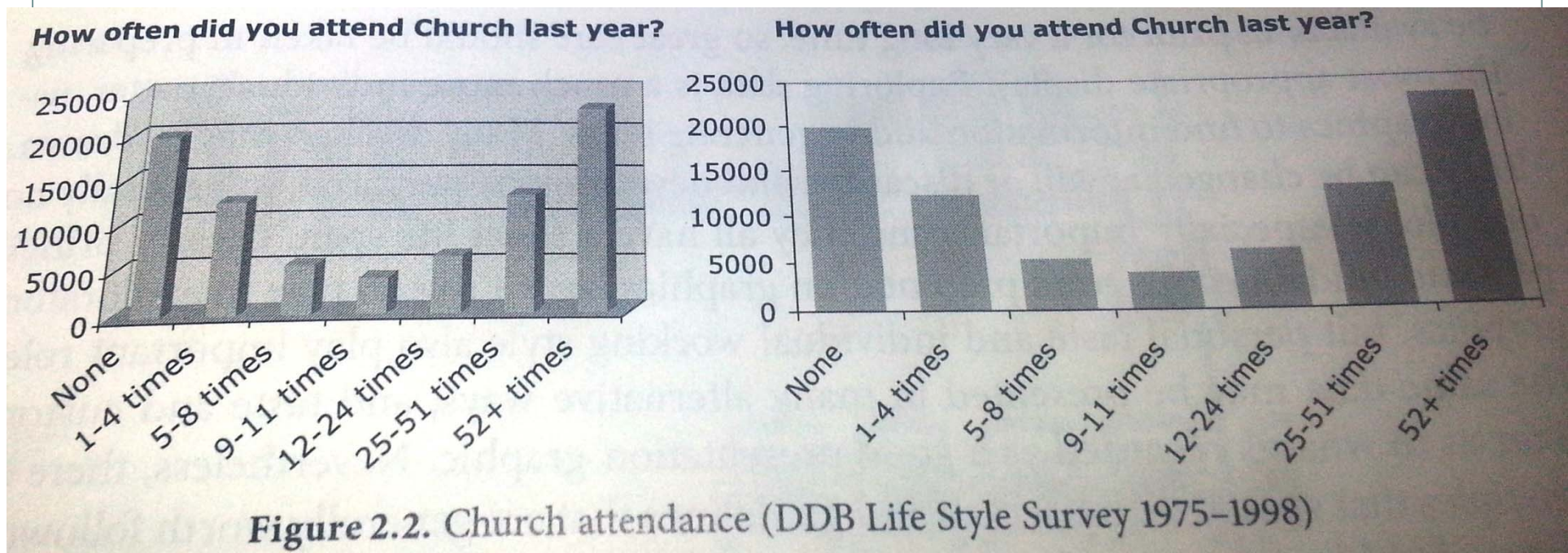
■ Apple ■ Blueberry ■ Key Lime ■ Pumpkin ■ Pecan



- How much ink for 5 numbers?
- People are bad at judging the relative magnitude of angles
- If you twist the rotation of the pie, you can cause observers to systematically misjudge the proportions
- The third dimension makes judgment worse

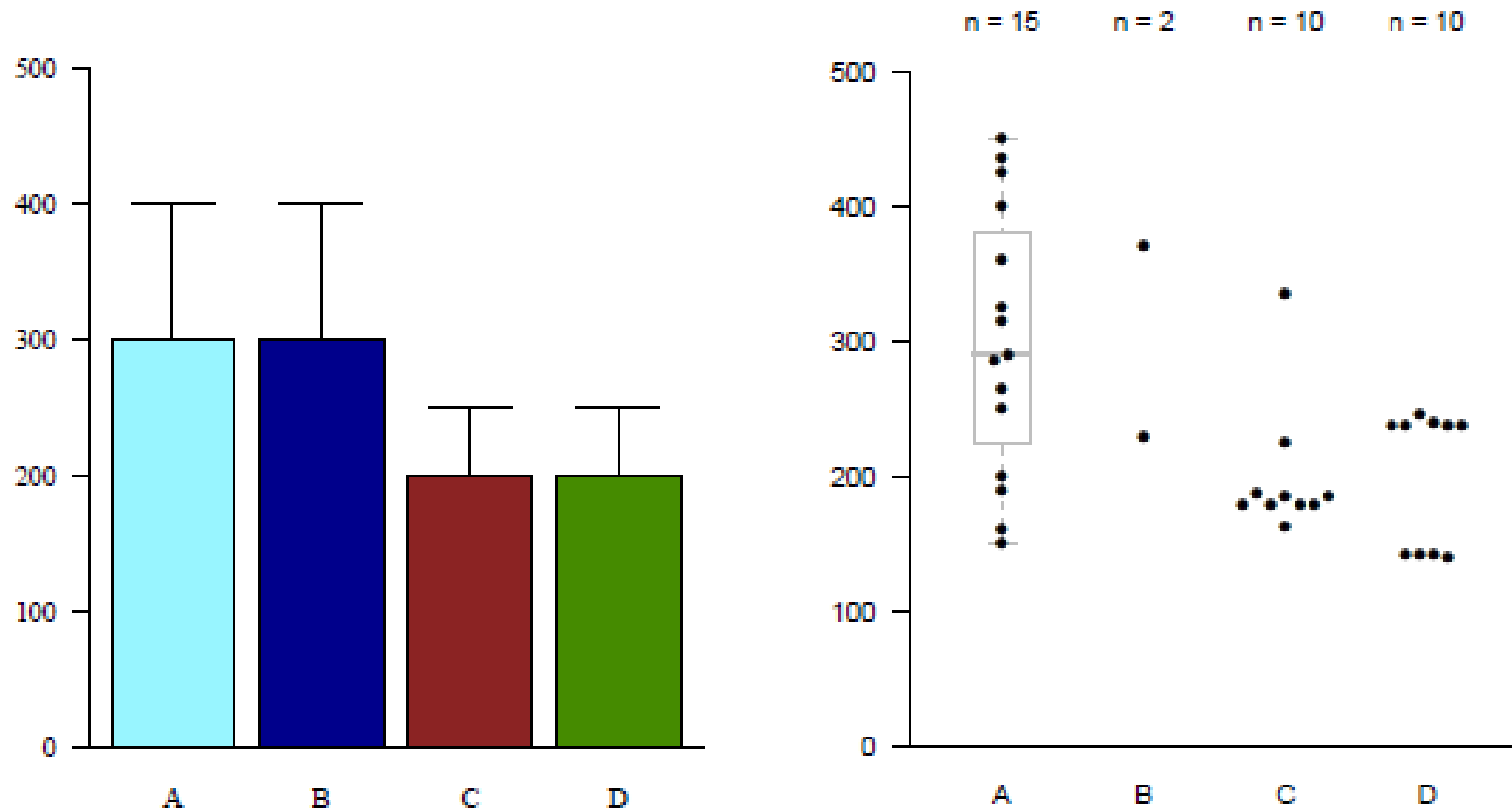
3D representation of 2D data

- 3D columns can cross gridlines
- Display of zero values would be misleading
- Difficult to make fair comparisons

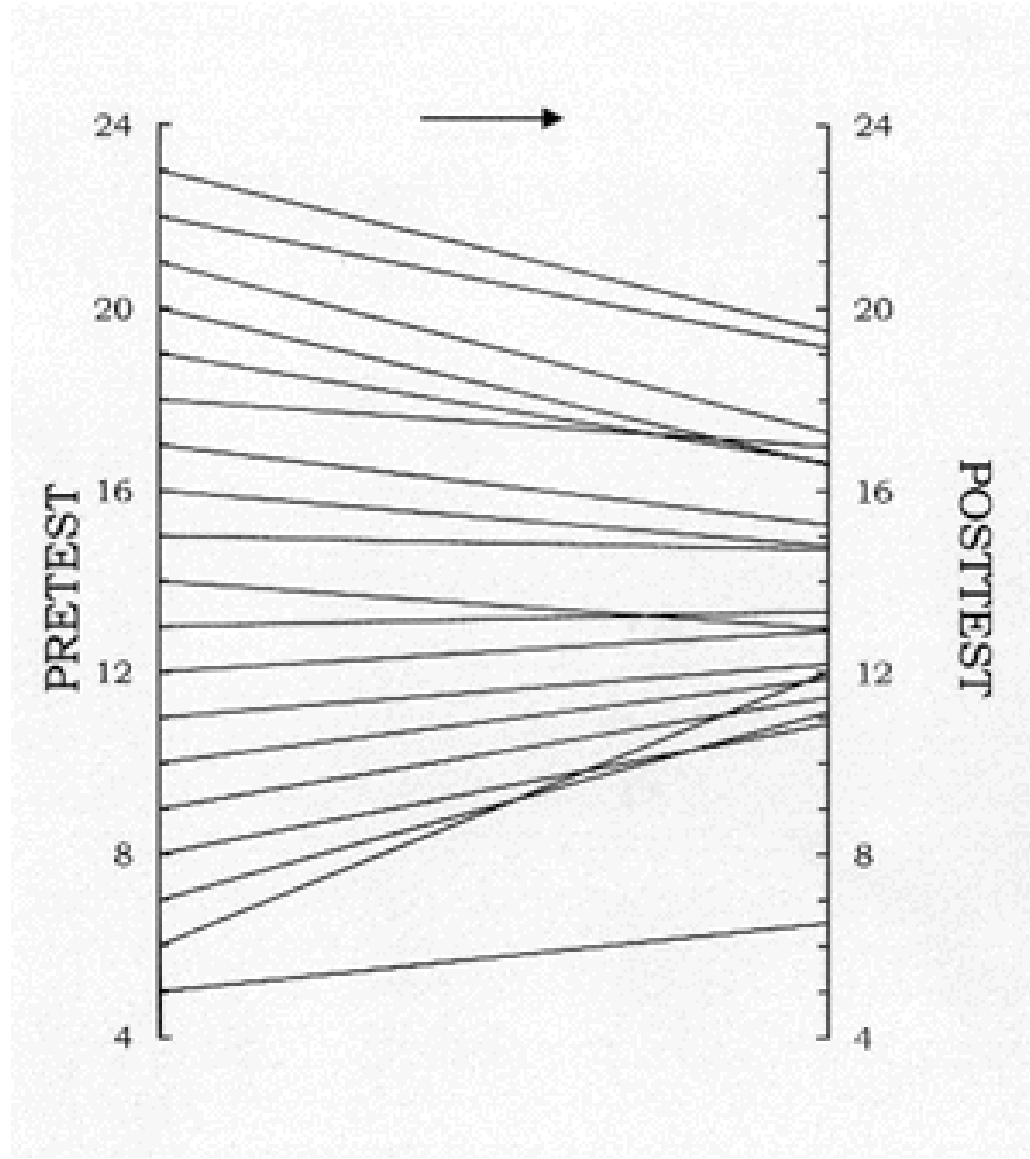


Dynamite Plots

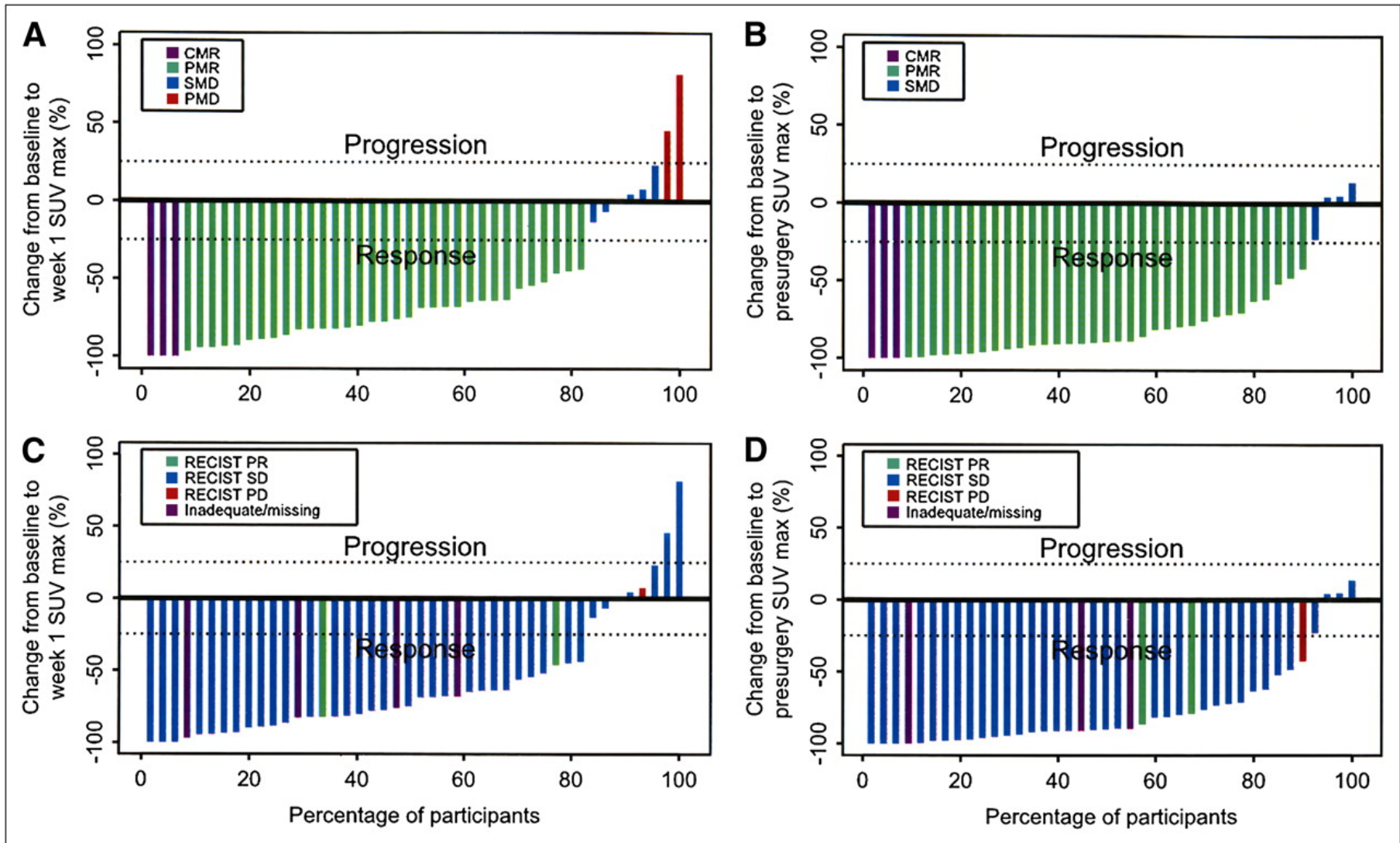
- You want to capture measure of central tendency and variability, but not like this...



Regression to the mean



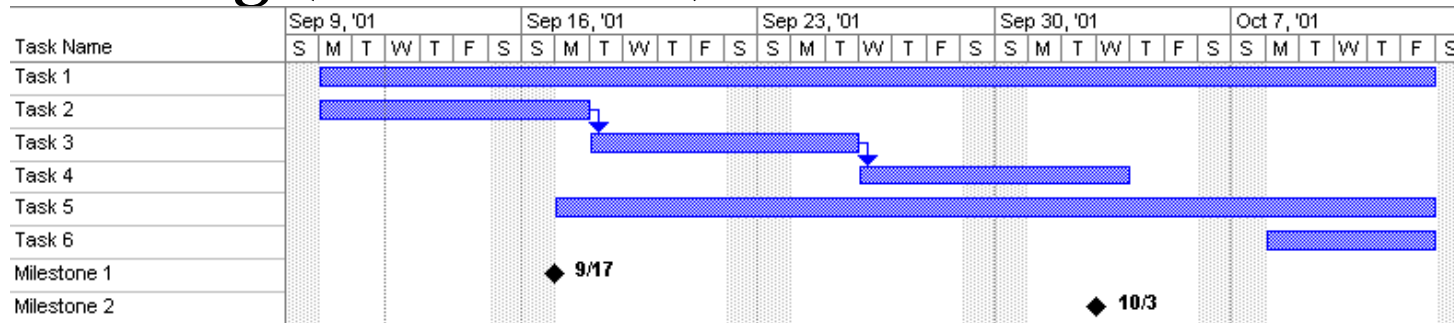
Waterfall Plot



Van den Abbeele et al. "ACRIN 6665/RTOG 0132 phase II trial of neoadjuvant imatinib mesylate for operable malignant GIST: monitoring with 18F-FDG PET and correlation with genotype and GLUT4 expression." *Journal of Nuclear Medicine* 53.4 (2012): 567-574.

Graphics for Grant Applications

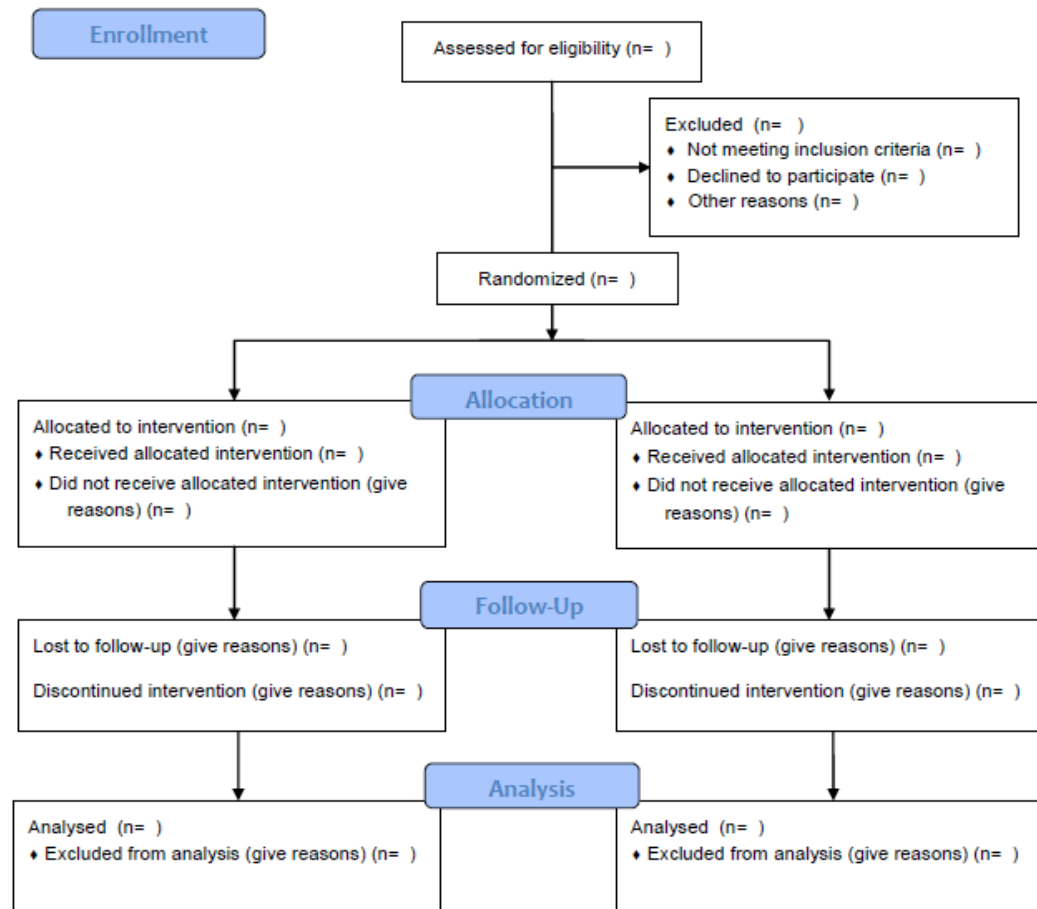
- Planning (Gantt Chart)



- Significance through Ngrams ([click here](#))
- Sample size and power analyses
 - Precision
 - Correlation
 - Two group comparison
 - Cluster randomized trial
- Make sure they are small with large, concise fonts and good resolution. Must stand on their own.

Study Flow Diagrams

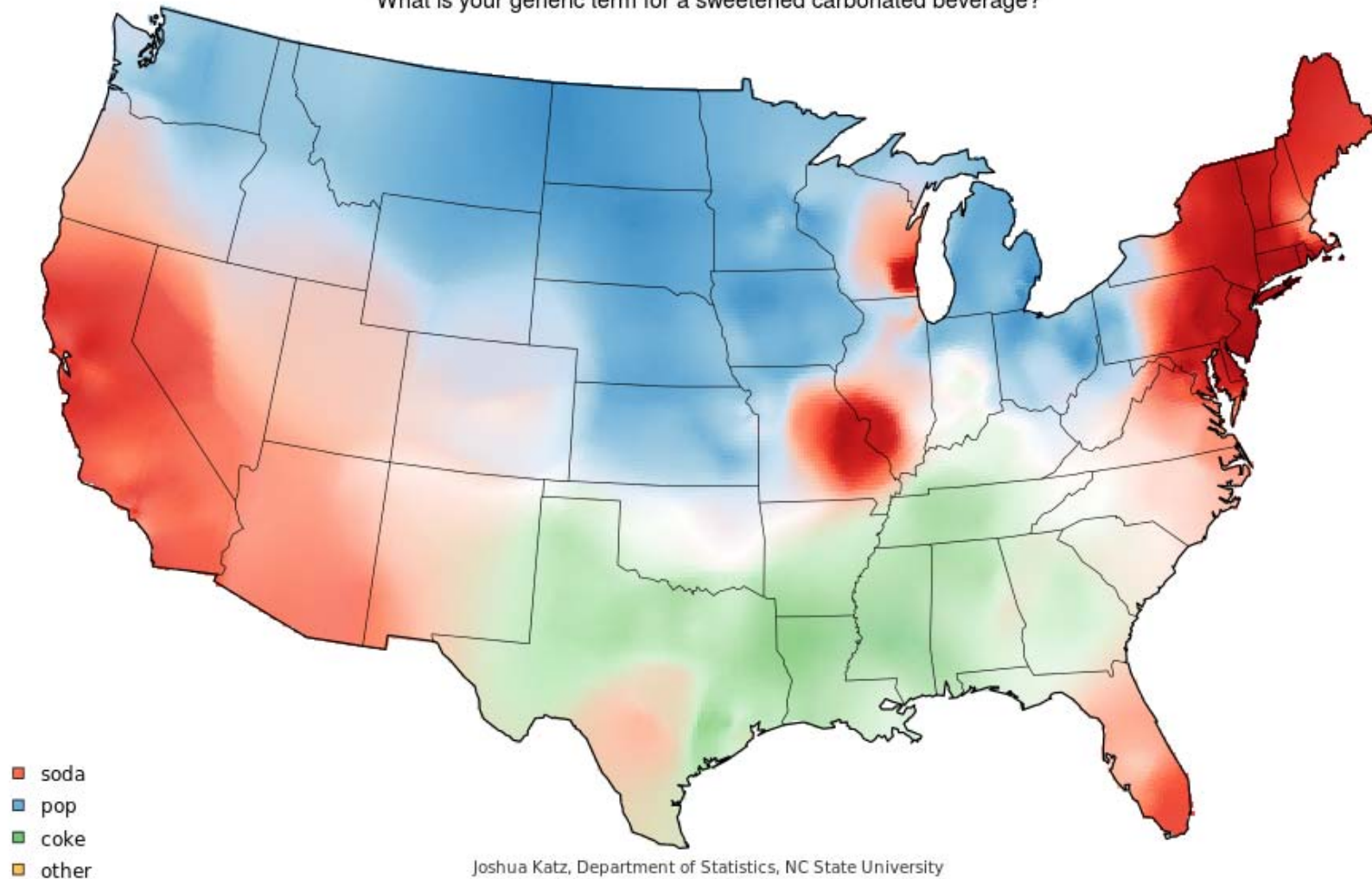
- CONSORT Flow Diagram
- STARD Flow Diagram
- STROBE Checklist and Flow Diagram (if applicable)



Dialect Map

Take the quiz: <http://www.nytimes.com/interactive/2013/12/20/sunday-review/dialect-quiz-map.html>

What is your generic term for a sweetened carbonated beverage?



Joshua Katz, Department of Statistics, NC State University
Based on data from the Harvard Dialect Survey by Bert Vaux & Scott Golder

Babies Dataset



The data provided here are part of the Child Health and Development Studies (CHDS)— a comprehensive investigation of all pregnancies that occurred between 1960 and 1967 among women in the Kaiser Foundation Health Plan (a prepaid medical care program). The women in the study were all those enrolled in the Kaiser Plan who had obtained prenatal care in the San Francisco–East Bay area and who delivered at any of the Kaiser hospitals in Northern California. The R dataset has a subset of this information collected for 1236 babies.

Baby boys born during one year of the study who lived at least 28 days and who were single births (i.e., not one of a twin or triplet).



Example dAtaViZ flowchart



Name	Role	Label	Units	Type	Values
GROUP	Predictor	Treatment		Binary	1 = Placebo; 2 = Treatment
AGE	Predictor	Age	Years	Continuous	18 - 75
SEX	Predictor	Gender		Binary	1 = Female; 2 = Male
HT	Predictor	Height	in.	Continuous	48 - 96
WT	Predictor	Weight	lbs.	Continuous	75 - 350
HCT	Predictor	Heart rate	beats/min.	Continuous	30 - 50
BPSYS	Predictor	Systolic BP	mmHg	Continuous	100 - 160
BPDIAS	Predictor	Diastolic BP	mmHg	Continuous	80 - 150
STAGE	Predictor	WHO stage		Discrete numeric	1 - 4
RACE	Predictor	Race		Categorical	1 = White; 2 = Black; 3 = Other
DATE1	Additional	Date of last visit			mm/dd/yyyy
COMPLIC	Outcome	Complications?		Binary	0 = No; 1 = Yes

Title



Nomogram

- Text

One Continuous Variable

- Histogram
- Boxplot
- Violin Plot
- Stripchart
- Density plot

One categorical variable

- Text

Layering to improve information

- = data-ink ÷