

Data Visualization

B. Connor McClellan

Graphics should

1. Show the data
2. Induce the viewer to think about the substance rather than about methodology, graphic design, etc.
3. Avoid distorting what the data have to say
4. Present many numbers in a small space
5. Make large data sets coherent
6. Encourage the eye to compare different pieces of data
7. Reveal the data at several levels of detail, from a broad overview to the fine structure
8. Serve a reasonably clear purpose: description, exploration, tabulation, or decoration
9. Be closely integrated with the statistical and verbal descriptions of a data set

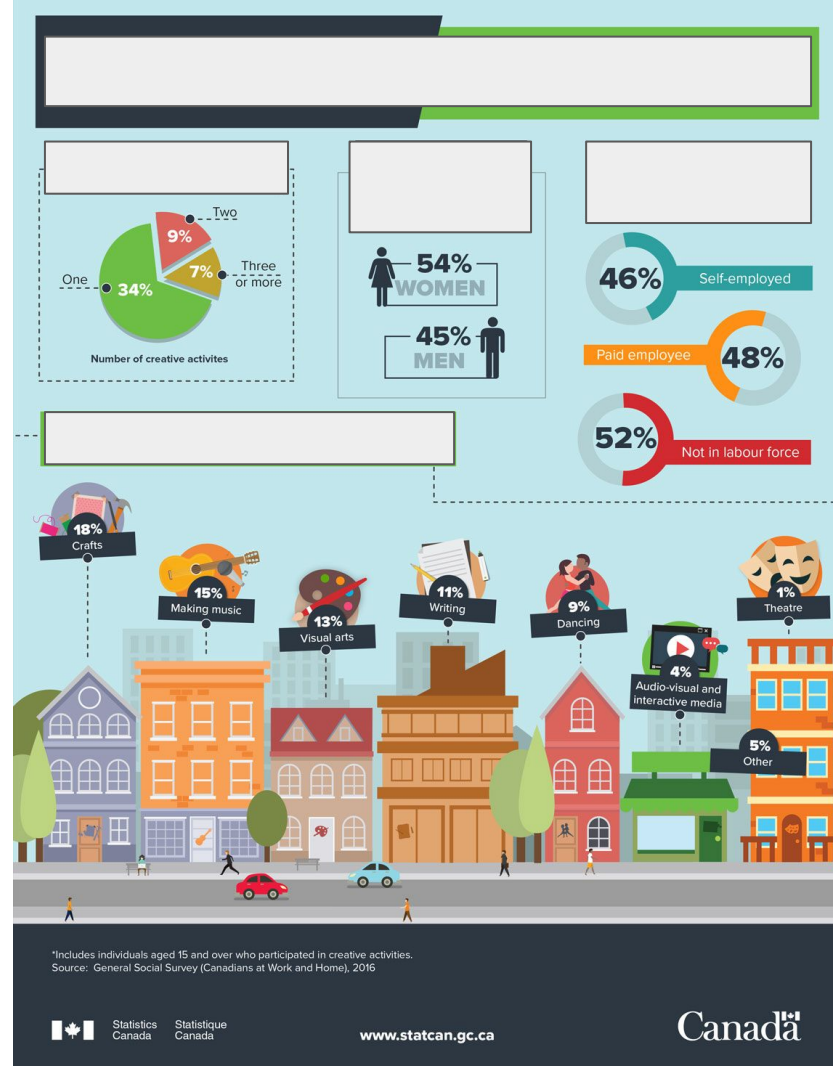
Graphics should

1. Show the data
2. Induce the viewer to think about the substance rather than about methodology, graphic design, etc.
3. Avoid distorting what the data have to say
4. Present many numbers in a small space
5. Make large data sets coherent
6. Encourage the eye to compare different pieces of data
7. Reveal the data at several levels of detail, from a broad overview to the fine structure
8. Serve a reasonably clear purpose: description, exploration, tabulation, or decoration
9. Be closely integrated with the statistical and verbal descriptions of a data set

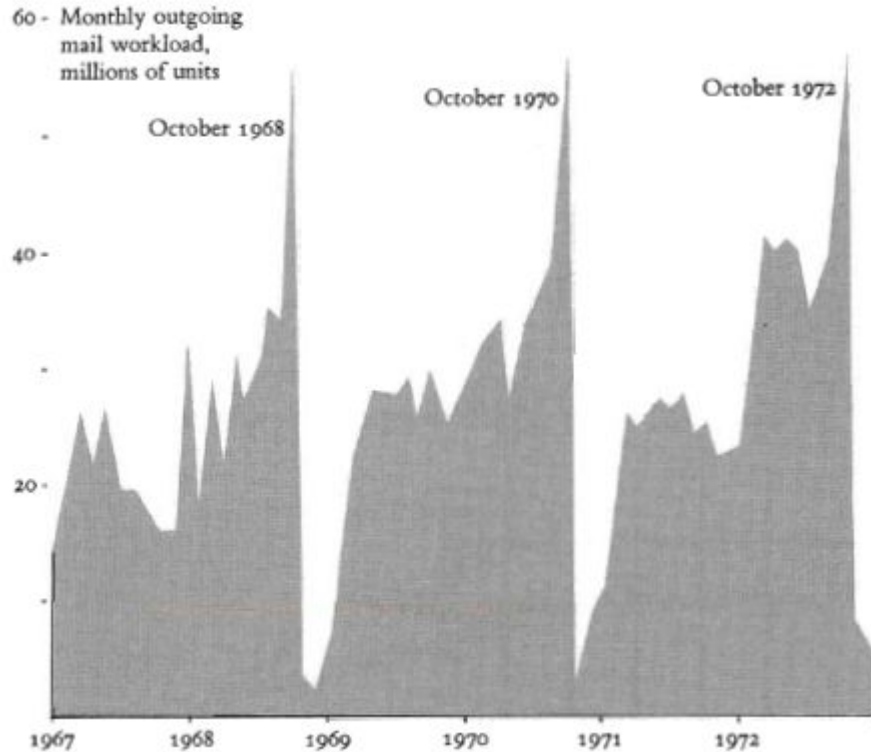
1. Show the data (Bad)

Here I've removed the descriptive titles. I challenge you to reach a single informative conclusion based on what is shown.

What's done wrong? There is an appalling lack of data actually present in this "infographic". The data should more or less speak for itself, even when the descriptions are removed.



1. Show the data (Good)



In contrast....

A question for you: what does this graphic show?

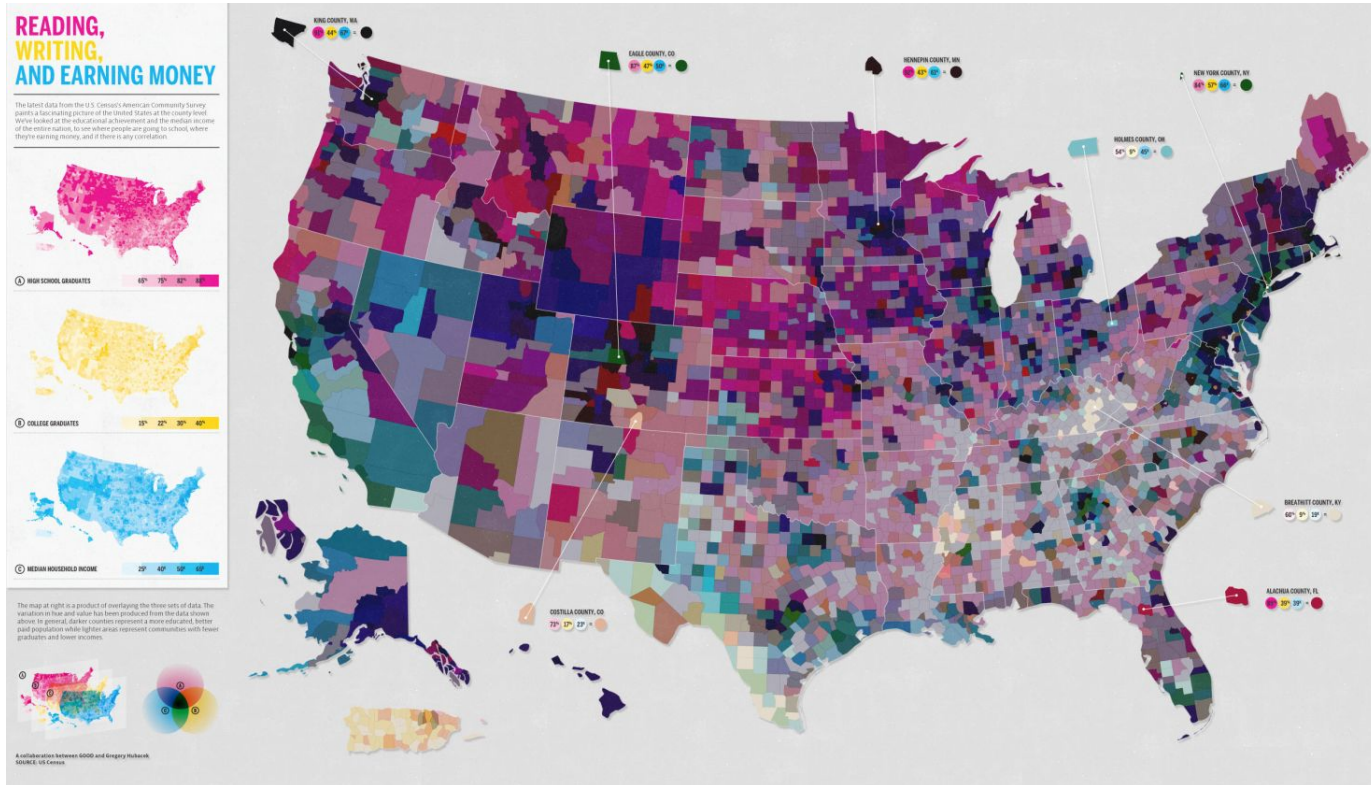
Answer: the outgoing mail of the U.S. House of Representatives peaks every two years, just before election day.

What's done right? The graphic *reveals the data* with very little additional text required. The date markings, "October 19XX", add just the slightest bit of context to let the reader conclude for themselves that the trend is caused by the election cycle.

Graphics should

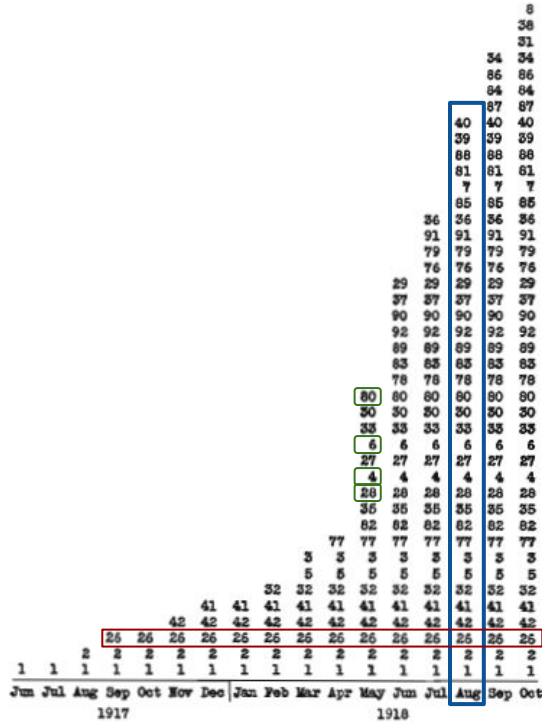
1. Show the data
2. Induce the viewer to think about the substance rather than about methodology, graphic design, etc.
3. Avoid distorting what the data have to say
4. Present many numbers in a small space
5. Make large data sets coherent
6. Encourage the eye to compare different pieces of data
7. Reveal the data at several levels of detail, from a broad overview to the fine structure
8. Serve a reasonably clear purpose: description, exploration, tabulation, or decoration
9. Be closely integrated with the statistical and verbal descriptions of a data set

2. Induce the viewer to think about the substance (Bad)



What's done wrong? The graphic would have been more informative as three separated colored graphics. Combining them *removes* information from the figure.

2. Induce the viewer to think about the substance (Good)



Leonard P. Ayres, *The War with Germany* (Washington, D.C., 1919), p. 102.

Shown: American divisions deployed in France as a function of time

But also...

The total number of divisions in France for each month

What particular divisions were in France in each month

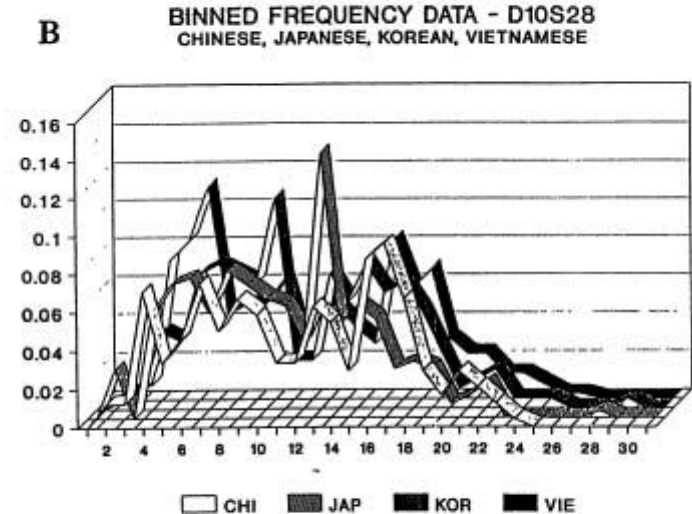
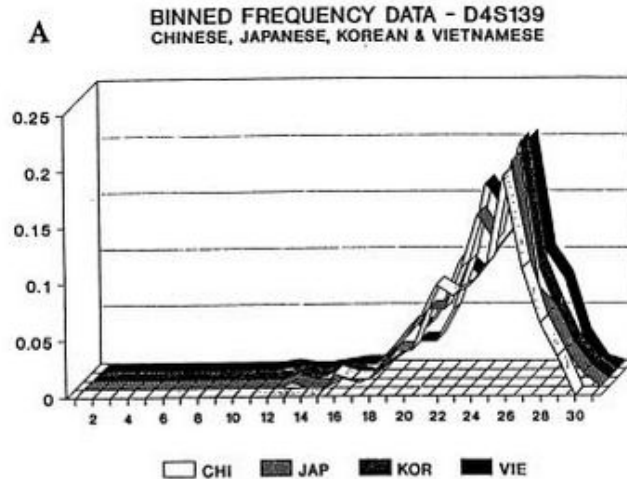
The duration of each division's presence in France

What's done right? The graphic has been reduced to such a simple, yet effective, format that multiple distinct conclusions can be drawn from a single plot. The data are revealed incredibly efficiently, maximizing the substance-per-ink.

Graphics should

1. Show the data
2. Induce the viewer to think about the substance rather than about methodology, graphic design, etc.
3. Avoid distorting what the data have to say
4. Present many numbers in a small space
5. Make large data sets coherent
6. Encourage the eye to compare different pieces of data
7. Reveal the data at several levels of detail, from a broad overview to the fine structure
8. Serve a reasonably clear purpose: description, exploration, tabulation, or decoration
9. Be closely integrated with the statistical and verbal descriptions of a data set

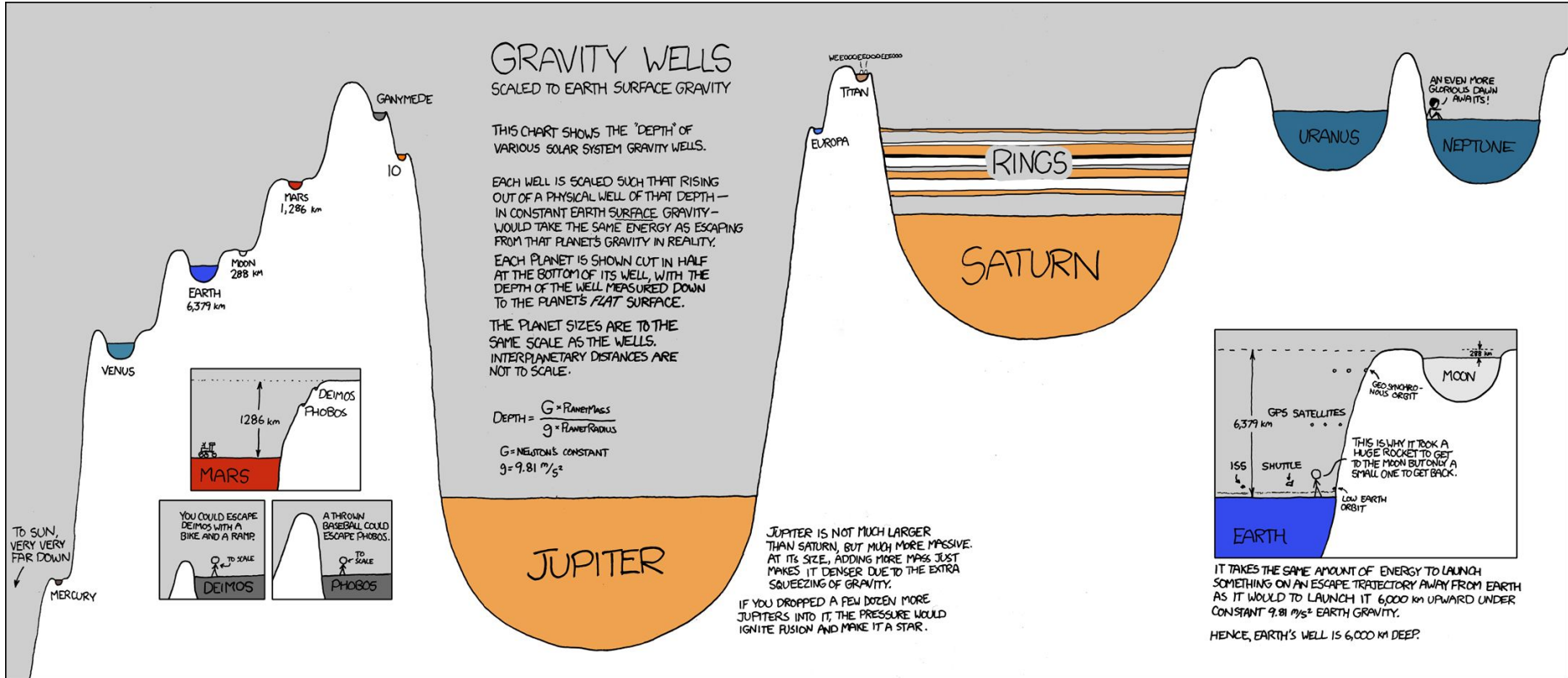
6. Encourage the eye to compare different pieces of data (Bad)



What's done wrong? The ribbon line style is completely unnecessary, and makes it nearly impossible to distinguish between the lines in the second panel. The eye is actively *discouraged* from seeing the data.

FIG. 4. Fixed bin distribution (histogram) for two loci and four Asian subpopulations (used with permission from John Hartmann): the boundaries of the 30 bins (vertical axis) are determined by the FBI; these bins are not of equal length. Sample sizes (numbers of individuals) for Chinese, Japanese, Korean and Vietnamese are 103, 125, 93 and 215 for D4S139 and 120, 137, 100 and 193 for D10S28. The horizontal axis is the bin number; bins are not of equal length.

6. Encourage the eye to compare different pieces of data (Good)



6. Encourage the eye to compare different pieces of data (Good)

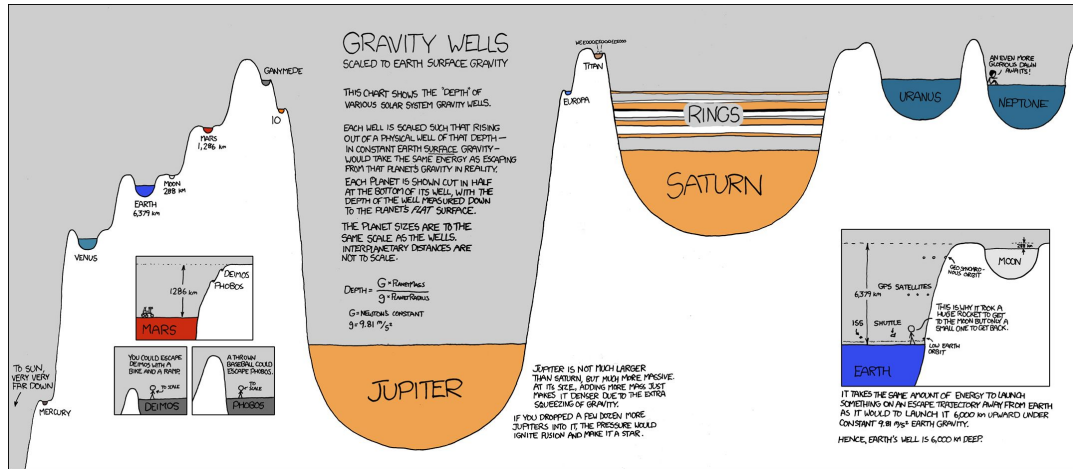
What's done right? Without reading a single word, you can tell the graphic conveys visual information regarding relative scales. Your eye is already comparing the data, whether intentionally or not!

Bonus:

Descriptive numbers are included where necessary

A complicated idea is shown in an uncomplicated fashion

No excess visual clutter



Aside: Table Design

Sometimes a table really is the best way to show data --- a graph may make a misrepresentation no matter how you slice it.

Remove

to improve
the **data tables** edition

How to make a good table

1. No gridlines, fills, bold, or other distracting stylistic choices
2. Left align text, right align numbers
3. Align headers with data
4. Whitespace creates a natural separation
5. Consistent precision, round numbers
6. Appropriate column sizing

Instrument	Star	k_{λ}	$\sigma_{k_{\lambda}}$
ARCTIC	Ref1	0.1338	0.0049
	Ref2	-0.0549	0.0128
	Ref3	-0.0482	0.0393
	Ref4	0.1312	0.0706
ARCSAT	Ref2	0.0335	0.0023
	Ref3	0.0418	0.0020
	Ref5	0.0329	0.0024
	Ref7	0.0416	0.0021

Table 1: Star-by-star atmospheric extinction coefficients and their uncertainties.

Summary:

- Show the data
- Induce the viewer to think about the substance rather than about methodology, graphic design, etc.
- Encourage the eye to compare different pieces of data

References:

www.darkhorseanalytics.com

www.xkcd.com

www.thesocietypages.org

The Visual Display of Quantitative Information, Tufte (2001)

Ayers, Leonard P, *The War with Germany*. Washington, D.C., 1919, p. 102

Roeder K (1994) DNA fingerprinting: A review of the controversy (with discussion). *Statistical Science* 9:222-278, Figure 4

Above all else show the data.

Maximize the data-ink ratio.

Erase non-data-ink.

Erase redundant data-ink.

Revise and edit.

