

# Visualizing quantitative data with R and RStudio

ME 447/547 Visualizing Data

---

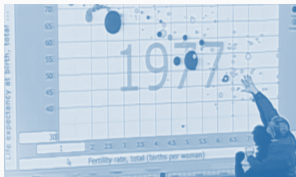
Richard Layton

December 2018

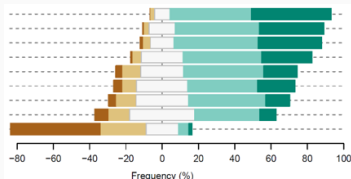
Rose-Hulman Institute of Technology

# The course is designed to develop your skills in three areas

Rhetoric



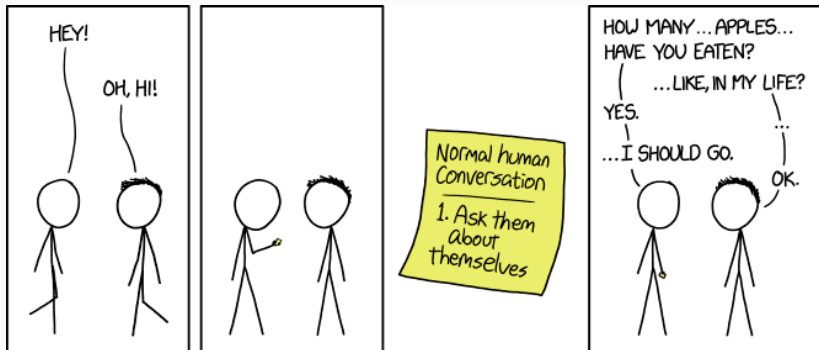
Repertoire



Means



# Please sit with someone you don't know and introduce yourself

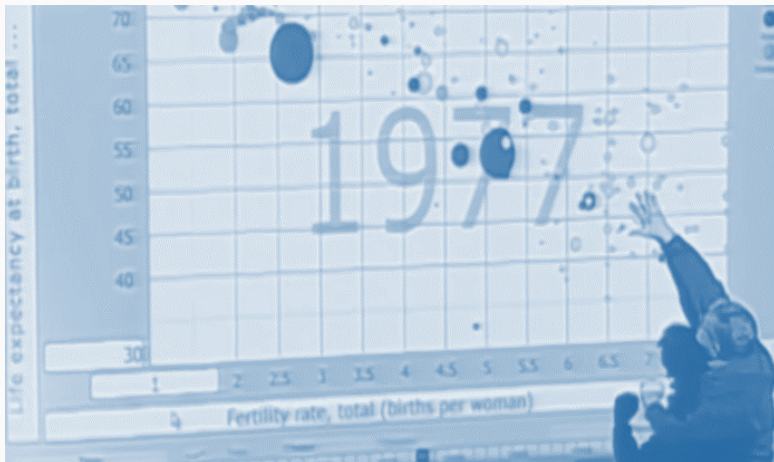


<https://www.xkcd.com/1976/>

# Visual rhetoric

---

# Designers shape information visually for rhetorical ends



Hans Rosling 2006 TED Talk

# Consider the argument

How did Hans shape the information visually?

What were his rhetorical goals?

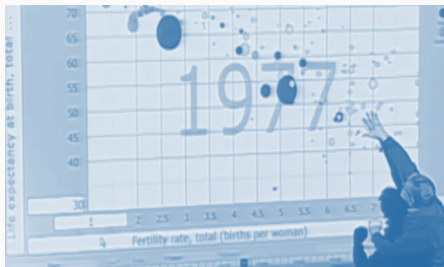


Image: TED2006

## Consider a different, less credible, visual argument

True or False:  $N_{\text{people on welfare}} > N_{\text{people with a full time job}}$

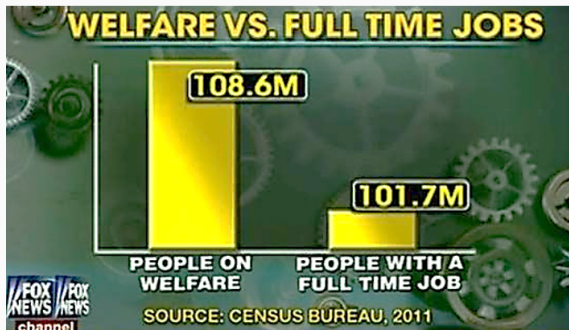


Image: Media Matters

## Consider a different, less credible, visual argument

True or False:  $N_{\text{people on welfare}} > N_{\text{people with a full time job}}$

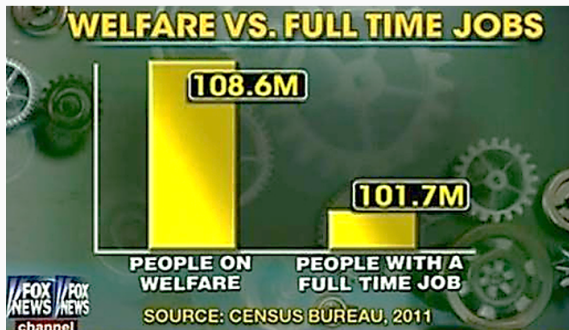


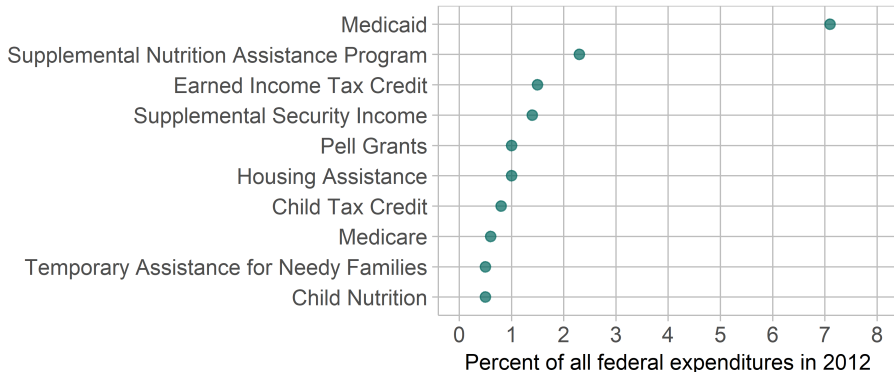
Image: Media Matters

False. One count is artificially high; the other is artificially low.  
The counts use different definitions of “people”.



# To avoid ambiguity, let's define "welfare"

## Federal means-tested programs and tax credits



In total, 17% of the 2012 US federal budget (\$590 B / \$3540 B).

Also, the **visual argument** belies the verbal argument

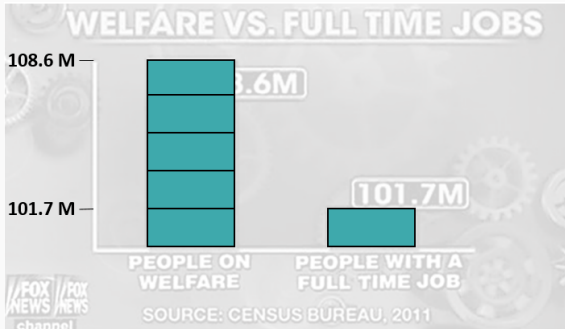
What is the **visual lie**?



# A visual argument prevails—as the designer well knows

Verbal argument: one number is 7% larger than the other

Visual argument: one group is **5 times larger** than the other



What were the designer's rhetorical goals?

# Ethical obligations are inherent in graph design



In data visualization, journalism meets engineering — Alberto Cairo

**journalism** increase knowledge among the public while minimizing harmful side effects

**engineering** give information a visual shape—model it, sculpt it—effectively and efficiently

(Cairo, 2014)

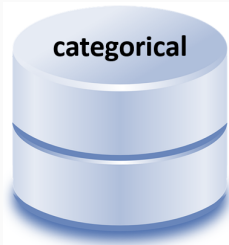
# Repertoire

---

# Graph design begins by understanding the **data structure** ...



Number of variables?  
Continuous or discrete?



Number of variables?  
Nominal or ordinal?  
Number of levels each?

... and by knowing the **prior art** suited to that structure

62

strip plot

box and whisker plot

multiway

scatterplot

dot plot

line graph

conditioning plot

63

scatterplot matrix

parallel coordinate plot

cycle plot

mosaic plot

financial (OHLC) plot

linked micromaps

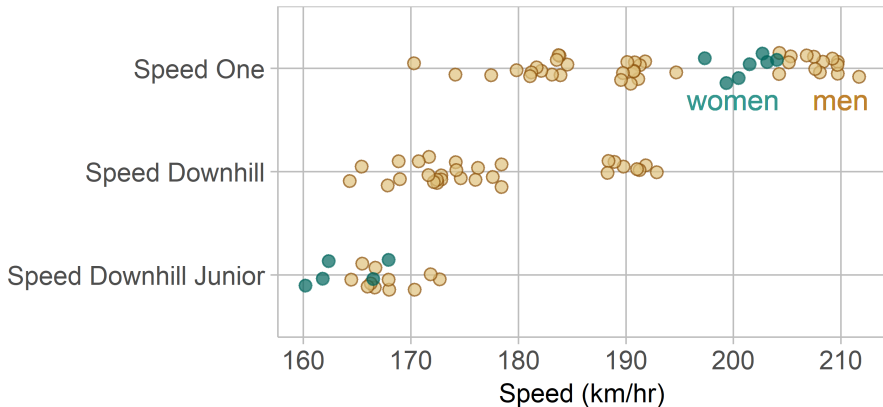
diverging stacked bar

## Gallery — strip plot, jitter plot, or 1D scatterplot

Quantitative variable: speed (continuous),  $N_{\text{obs}} = 91$

Categorical variables: event (3 levels), sex (2 levels), both nominal

2011 World speed-skiing championship



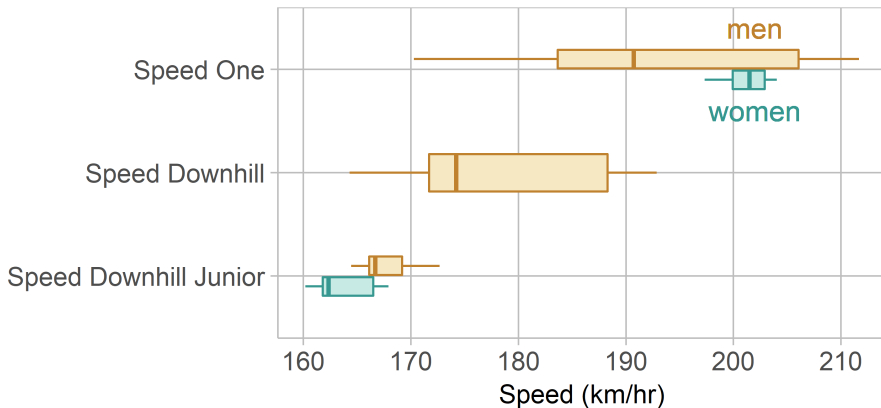


## Gallery — box and whisker or box plot

Quantitative variable: speed (continuous),  $N_{\text{obs}} = 91$

Categorical variables: event (3 levels), sex (2 levels), both nominal

2011 World speed-skiing championship



Population data by county (Robbins Figure 8.15)

Livestock (Cleveland Figure 6.1 and 6.2)

Any of the midfieldr data

Life expectancy by country (Kirk Figure 6.30)

Crime rates by state (Kirk Figure 6.31) bubble plot

Weight and height by sport (Unwin Figure 5.11)

Olive oil (Unwin Figure 10.12 and 10.14)

regression, linear and loess

## Gallery — Cleveland dot plot

State areas (Robbins Figure 4.3) with log base2 scale

midfieldr graphs, e.g., grad rate, starters, etc.

Brain and body mass by species (Cleveland/Elements Fig. 1.7)

Language speakers (Cleveland/Elements Fig. 1.9)

Fraction of journal space in graphs (Cleveland/Elements Figure 3.22)

## Gallery — line graph

Energy data (Robbins Figure 5.3)

Stock market (Robbins Figure 6.10)

Blood level data (Robbins Figure 7.21)

Car production (Robbins Figure 8.13)

Rubber properties (Cleveland Figure 4.4)

NOx (Cleveland Figure 4.6)

Solar radiation (Cleveland Figure 5.5)

## Gallery — scatterplot matrix

Energy data (Robbins Figure 5.6)

Blood level data (Robbins Figure 7.22)

Rubber properties (Cleveland Figure 4.1)

Solar radiation (Cleveland Figure 5.1)

Pima Indians diabetes (Unwin Figure 1.9)

Crime rates (Unwin Figure 5.12)

Nutrient contents (Kirk Figure 6.32)

Food data set (Unwin Figure 6.2)



St Louis Science Center attendance (Robbins Figure 4.18)

CO<sub>2</sub> (Robbins Figure 4.20) or (Cleveland Figure 3.75))

Arctic ice (?)

Monterrey Bay Aquarium (Robbins Figure 5.10)

Titanic (Unwin Figure 7.2)

Treatment (Unwin Figure 7.9)

## Gallery — financial (OHLC) plot

Dow Jones (Robbins Figure 5.16)

Gold price (Robbins Figure 8.2)

Soybean data (Robbins Figure 5.11)

Election funding raised (Kirk Figure 6.50)

Race and location (Kirk Figure 6.52)

Tornado data (Brunsdon Figure 5.1)

CATME world map

Broad Street cholera map (Bivand Figure 4.7)

## Gallery — diverging stacked bar

Literacy proficiency (Kirk Figure 6.24)

Student volunteers (Evergreen Figure 5.2)

My teaching evaluations

One from the Robbins paper

# Implications for the designer



Grasp the data structure



Explore using suitable designs



Refine the logic of your argument



Consider original designs if required by the story



Meet the needs of the audience



# Means

---

# Use the right tool for the job



RStudio

primary interface, integrates all our software



R

tidying data and creating graphs



R markdown

writing the portfolio, interleaving prose with code



Git

local version control




GitHub

collaborating and publishing the portfolio



# The main topical threads weave through the calendar

data  
software  
visual rhetoric  
repertoire of graphs  
portfolio

## calendar

 paper reprint, with permission

 e-copy on Moodle, with permission

w	d	agenda & assignments
1	M	Course goals and outcomes <a href="#">Sign-out two reprints</a>
	T	Introduction to visual rhetoric <a href="#">Install software</a>
	R	Relating data structure to graph design  Doumont (2009) Designing the graph
	F	<a href="#">Software lab</a>
2	M	Graph basics with ggplot2 Practice
	T	 Tufte (1997) Decision to launch Challenger
	R	Data basics Practice
	F	<a href="#">Data lab with file management</a> Return reprints

<https://github.com/DSR-RHIT/me447-visualizing-data>

# References

Cairo A (2014) *Ethical infographics*. The Investigative Reporters and Editors Journal, Spring 2014  
<https://www.dropbox.com/s/pqgm02yz0pgju4/EthicalInfographics.pdf>