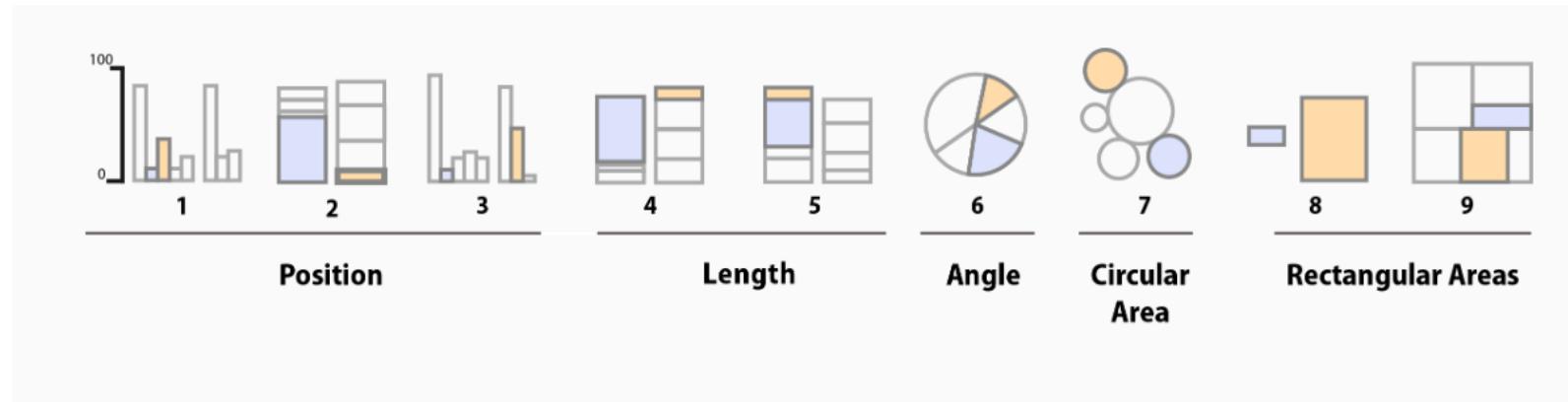


# Principles of data visualization wrap-up

5 February 2020

*Modern Research Methods*

# Last Time: Principles of visualization



Think of plotting as communication – you want to maximize the likelihood that your audience gets your message.

Form groups of ~4

Look at the plots on paper in front of you.

Discuss the effectiveness of each plot at communicating its message. What would you change?

[~ 5 min]

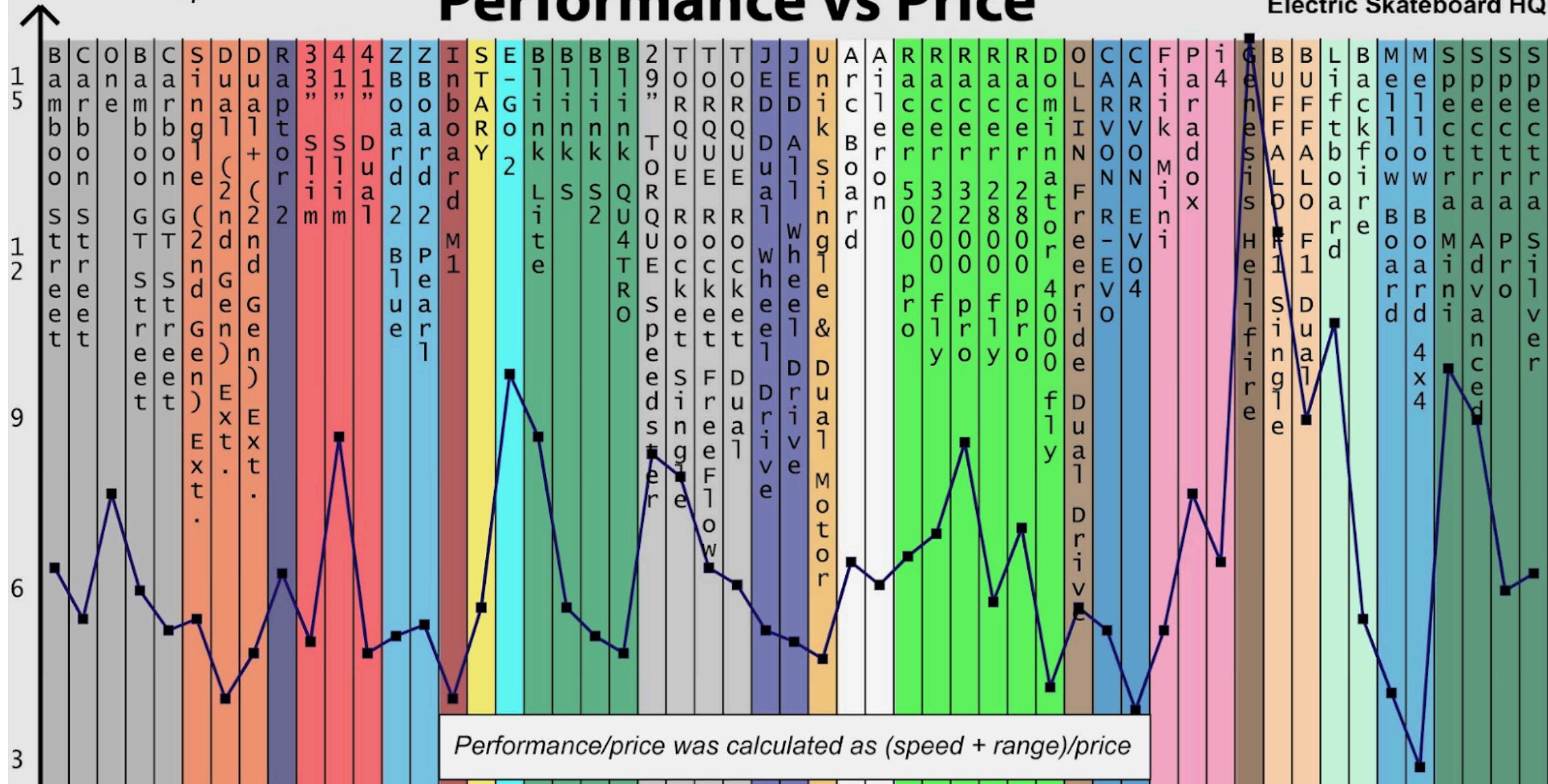
# ELECTRIC SKATEBOARDS (June 2017)

Brought to you by,



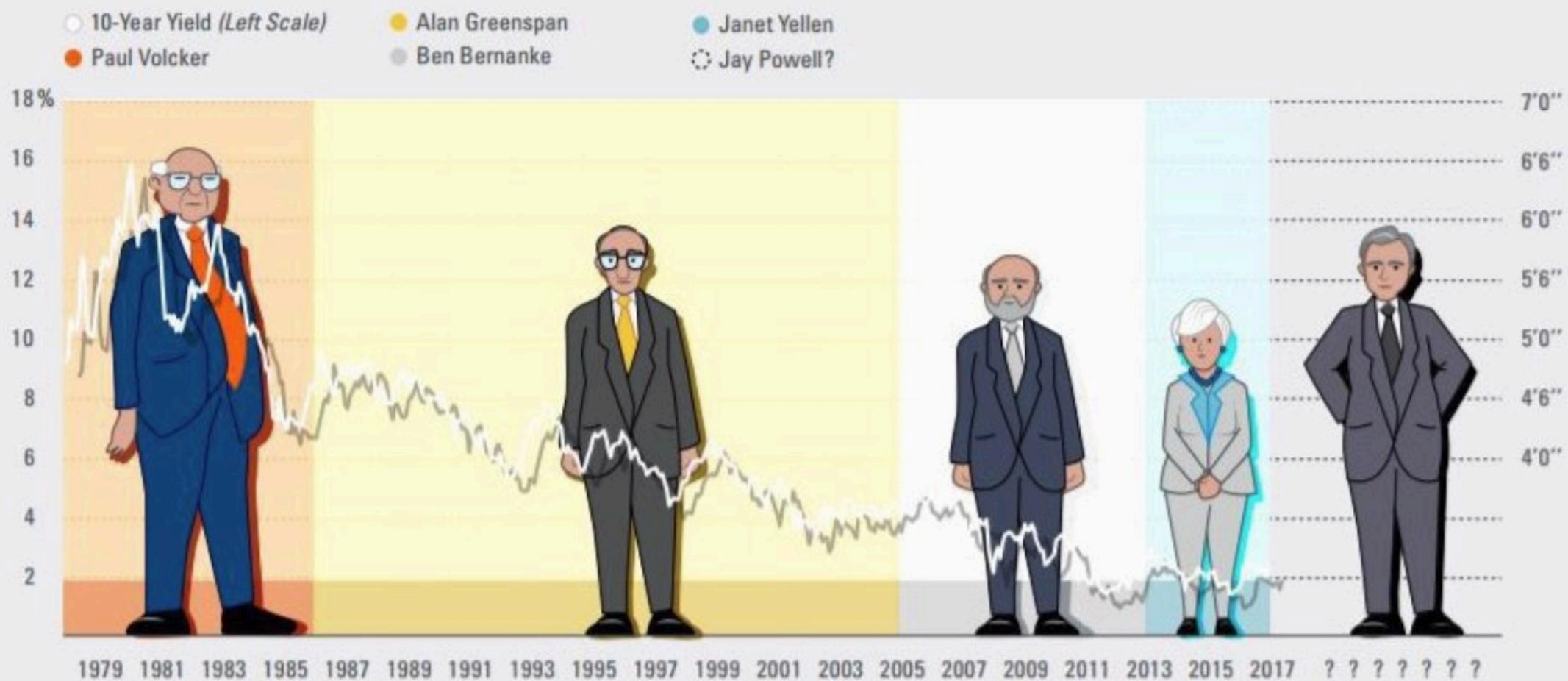
*Performance/price*

## Performance vs Price



2

## BOTH THE HEIGHT OF THE FED CHAIR AND RATES HAVE FALLEN OVER TIME, COULD A TALLER FED CHAIR MEAN RATES RISE?



Source: LPL Research, Bloomberg 10/22/17

We don't actually believe that interest rates are determined by the height of the Fed chair, but it has been an interesting coincidence.

## Over 100 Million Now Receiving Federal Welfare

2:40 PM, AUG 8, 2012 • BY DANIEL HALPER 



PRINT



EMAIL PAGE



LARGER TEXT



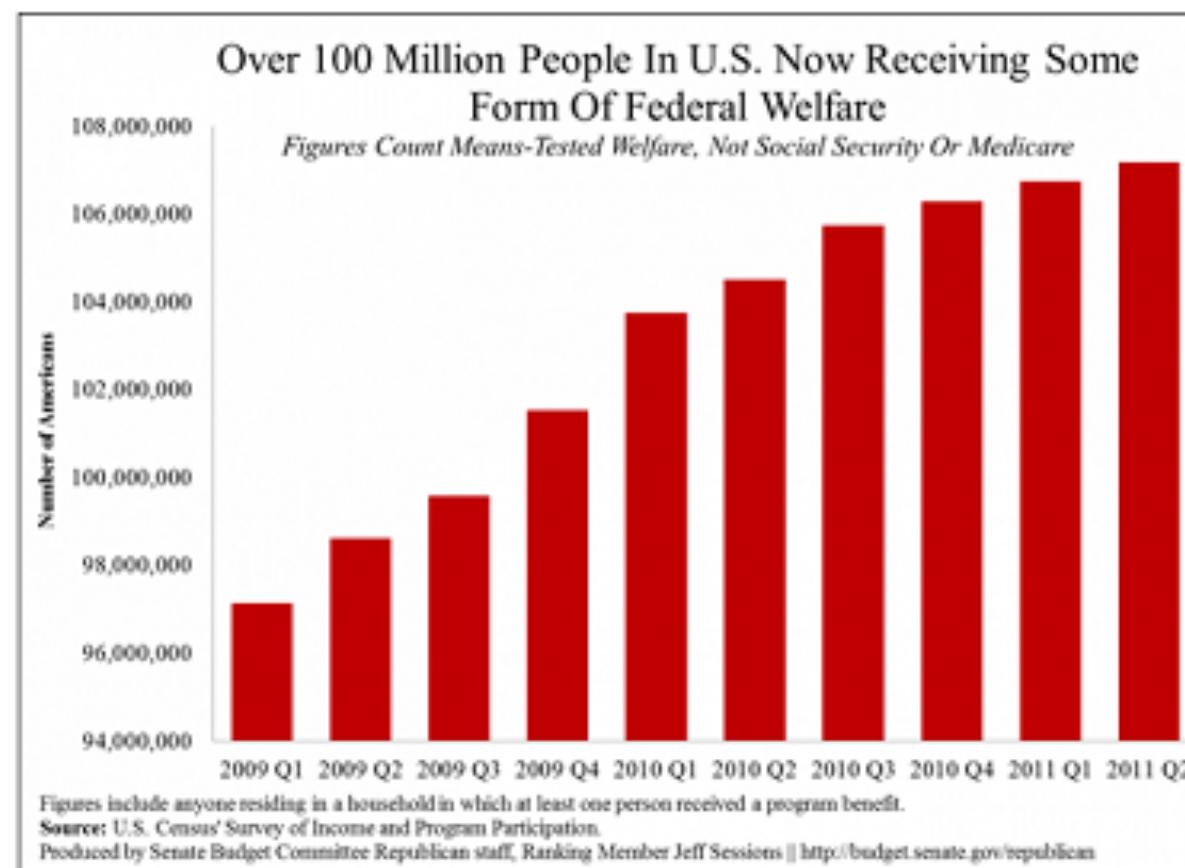
SMALLER TEXT



ALERTS

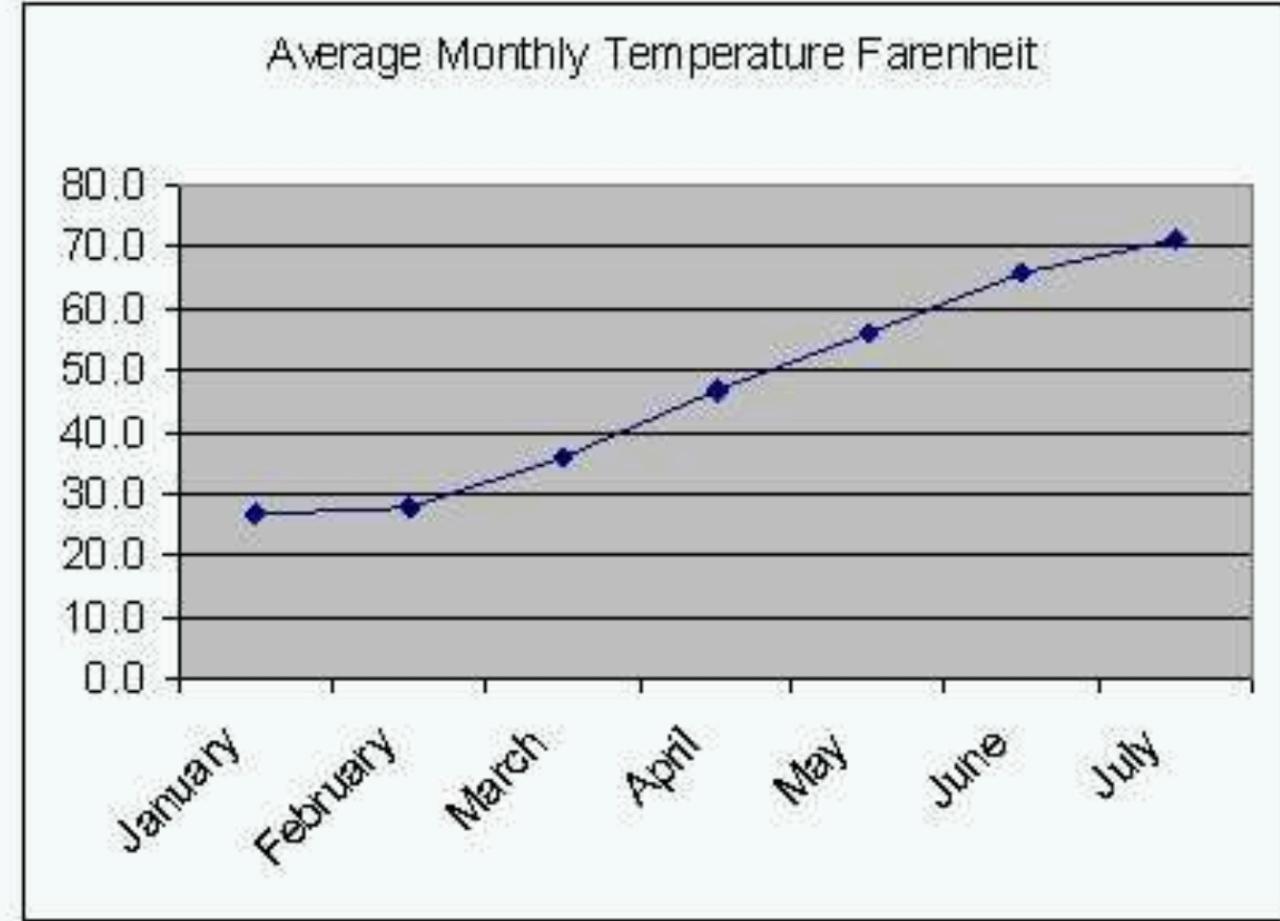


A new chart set to be released later today by the Republican side of the Senate Budget Committee details a startling statistic: "Over 100 Million People in U.S. Now Receiving Some Form Of Federal Welfare."



# Global Warming out of Control!

Average Monthly Temperature	
	Fahrenheit
January	26.4
February	28.0
March	35.8
April	46.6
May	56.3
June	65.8
July	71.2

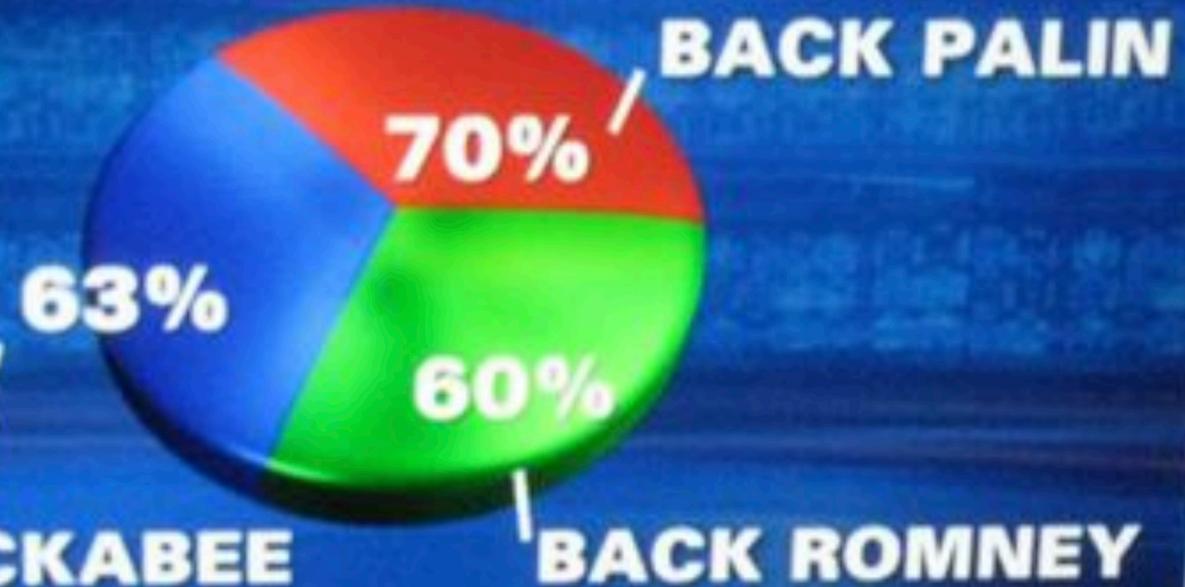


Average monthly temperature in New Haven, CT.

Figure 2

# 2012 PRESIDENTIAL RUN

GOP CANDIDATES



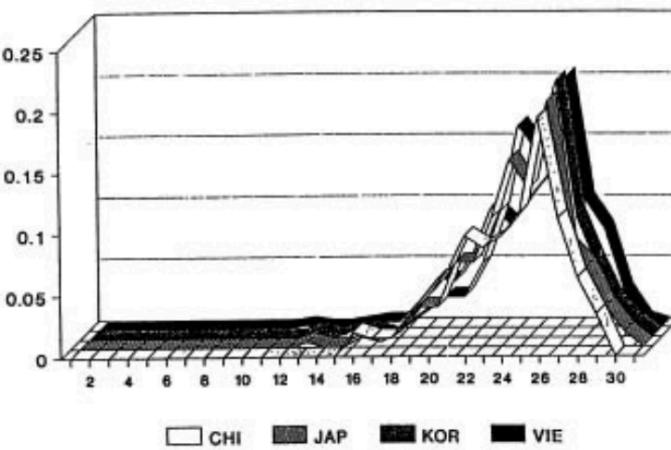
SOURCE: OPINIONS  
DYNAMIC

FOX

47°

**A**

BINNED FREQUENCY DATA - D4S139  
CHINESE, JAPANESE, KOREAN & VIETNAMESE

**B**

BINNED FREQUENCY DATA - D10S28  
CHINESE, JAPANESE, KOREAN, VIETNAMESE

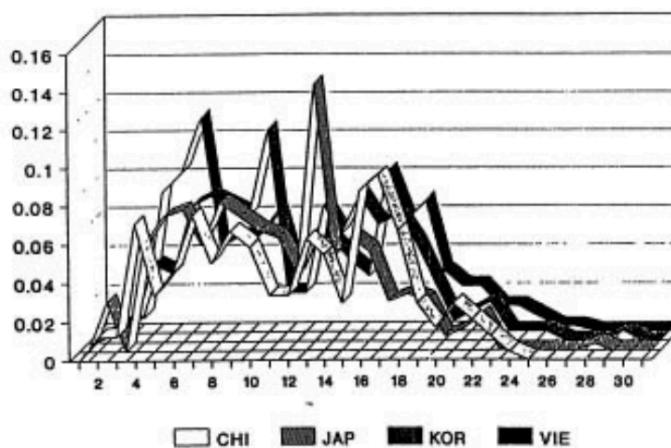


FIG. 4. Fixed bin distribution (histogram) for two loci and four Asian subpopulations (used with permission from John Hartmann): the boundaries of the 30 bins (vertical axis) are determined by the FBI; these bins are not of equal length. Sample sizes (numbers of individuals) for Chinese, Japanese, Korean 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.

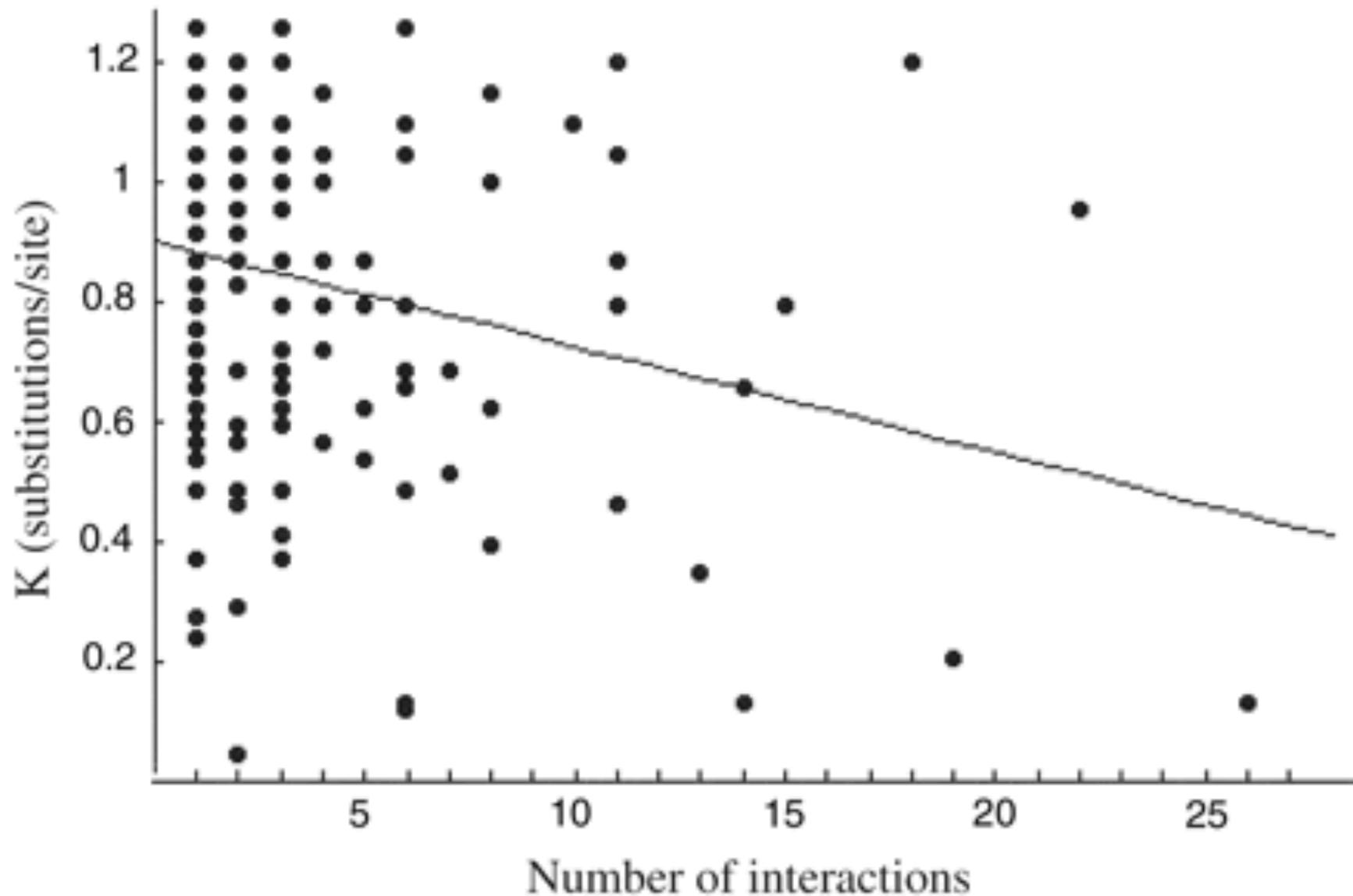
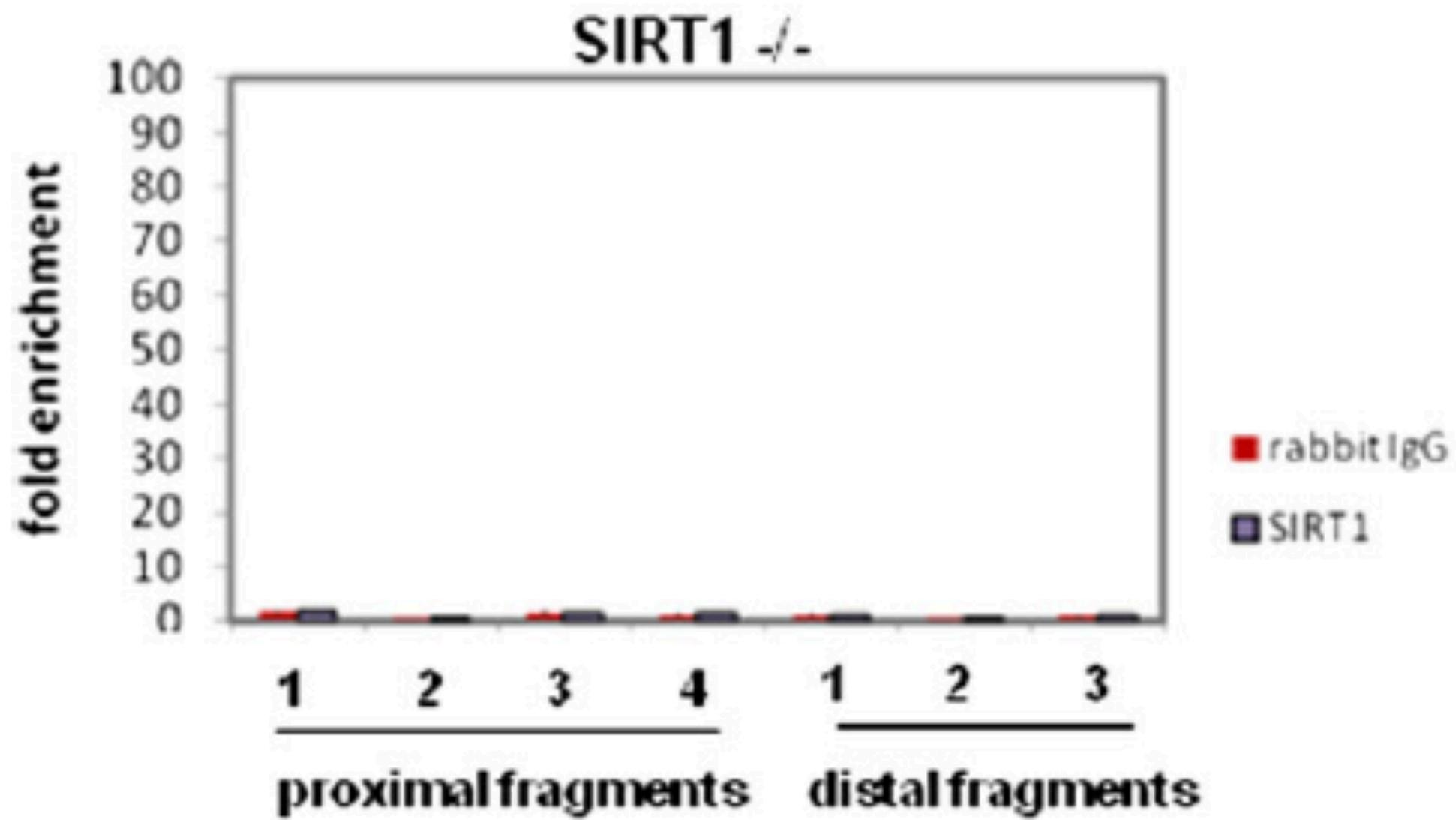
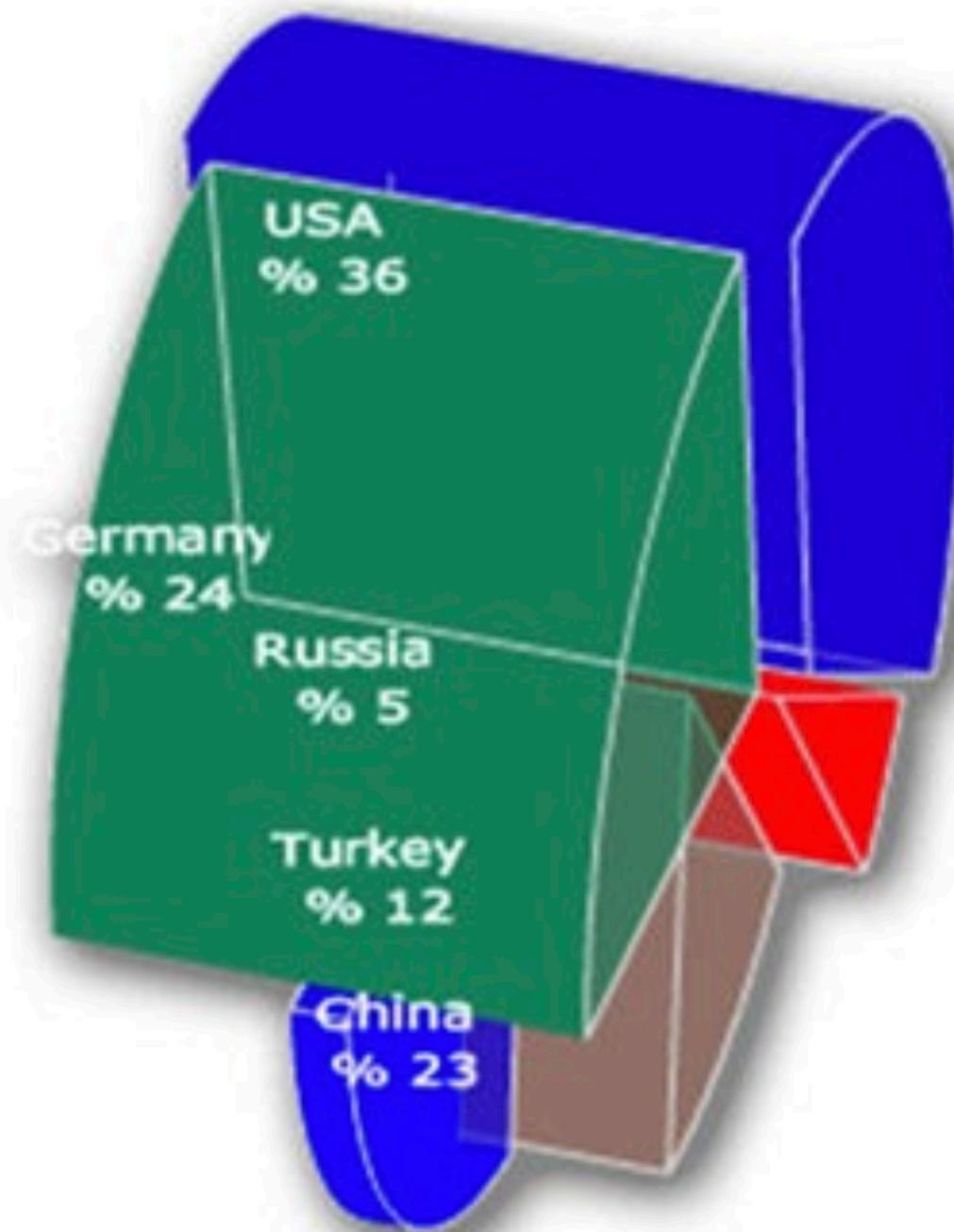


Figure 1

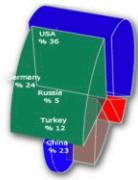
The relation between the number of protein-protein interactions ( $I$ ) in which a yeast protein participates and that protein's evolutionary rate, as estimated by the evolutionary distance ( $K$ ) to the protein's well-conserved ortholog in the nematode *C. elegans*.



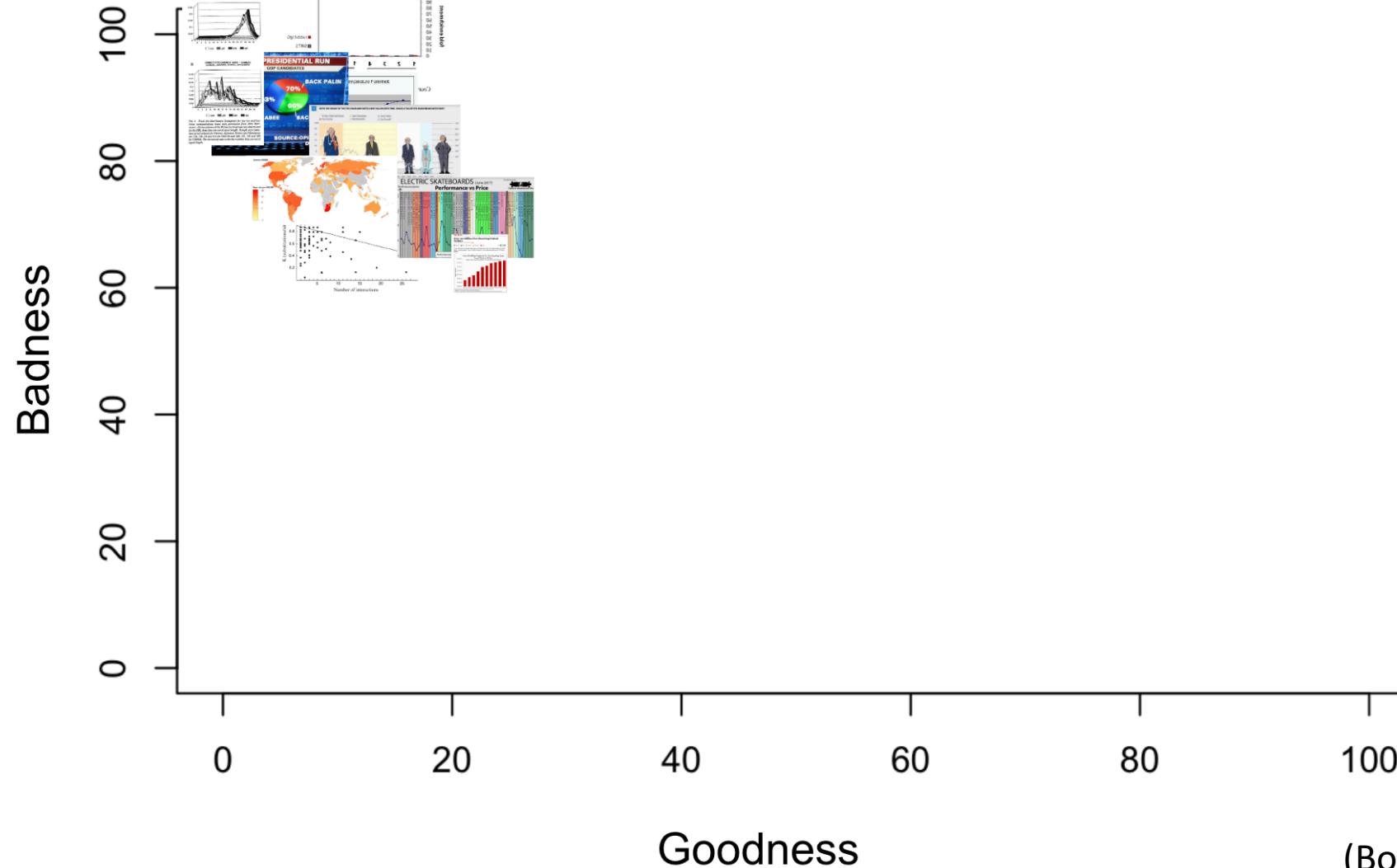
Courtesy: Donmez et.al Cell, 142, 320-332 (2010).



# Bonus plot



Badness over goodness of ten bad plots



(Bonus plot from Jaea)

# Assignment 1

- Is graded
- Nice work overall
- Didn't penalize for formatting issues this time, but will on Assignment 2
  - Make sure your code and text are formatted as we discussed in lab
  - Make sure you add plain text, where appropriate, to answer the question. "What is the smallest value of n?" "The smallest value of n is x."
  - If you need help with this, more than happy to help in office hours
- A few common errors...

## Exercise 7

Use Boolean operators to return only the rows that contain trials (rows): (a) with category as vegetables, (b) with category as animals and a trial number less than 7, (c) with category as vegetables or animals, (d) with at least one basic level response.

- **proportion\_basic\_level\_responses** - This is the variable that we measured. It refers to the proportion (out of 2 possible) of basic level examples that a participant selected.

d

```
lf_data %>%
  filter(proportion_basic_level_responses > 0, condition == "one")
```

OR

```
lf_data %>%
  filter(proportion_basic_level_responses >= 0.5, condition == "one")
```

### Exercise 8

The following code selects all trials (rows) where the condition was either “three\_subordinate” or “one.” Rewrite this code in a way that uses the `%in%` operator.

```
filter(lf_data, condition == "three_subordinate" | condition == "one")
```

```
filter(lf_data, condition %in% c("three_subordinate", "one"))
```

**Exercise 13**

On the first day of class, we talked about the “Sally Anne Task” that measures children’s understanding of theory of mind ([example videos](#)). Describe four variables that you could measure in this task to assess children’s theory of mind performance. Specifically, describe (1) one qualitative variable, (2) one quantitative - binary variable, (3) one quantitative - numeric, and (4) one quantitative - real variable. For each variable, give a one sentence description of the variable, AND one example value of that variable with units.

- Many people gave examples of variables that were of the right type, but didn’t assess performance.

1. Qualitative: Evaluate the accuracy of the child's reasoning/explanation behind their choice (not just the choice itself). Example: Sally would think the block is in the box because that's where she put it before she left the house. (no units, just a qualitative description)
2. Binary: Whether or not the child answered correctly as to where Sally would think the block is. Example: 0 = correct; 1 = incorrect
3. Numeric: If the task was repeated with different scenarios per child, determine the percent of times the children answered correctly (to make sure it's not based on chance; that they're not simply guessing). Example: answered correctly 80% of the time
4. Real: Calculate the amount of time spent thinking about the situation before the child provides an answer. Example: thinking time = 2 minutes, 15 seconds

(Shruti Murali)

Qualitative: When explaining why they think the box is where it is, does the child use first person pronouns, third person pronouns, none at all or a mixture (i.e. "the box is in there"?). This might show to what extent they are relating to the characters and seeing their own experience as separate. Example: for subject 1, pronouns = "first\_person". Quantitative binary: Does the child 'correctly' identify where Sally thinks the box is? Example: correct\_location = 0 if incorrect, 1 if correct. Quantitative numeric: How many times does the child use words like "think", "thought", "knew", or "remember" in relationship to the characters? This might show whether they are thinking about other people's thought processes. Example: used these words 3 times. Quantitative real: How long does the child spend explaining their answer? This could show how hard the child has to think about the question and how well they understand the situation. Example: a 34.5 second response.

(Isobel Stephen)