

Elegant Data Visualization with ggplot2

Lesson 2

API 209: Advanced Quantative Methods

TF: Rony Rodriguez-Ramirez

Summer 2024

Plan for today

1. List
2. List
3. List
4. List

First Part

Just show me the data!

```
head(my_data, 10)
```

```
## # A tibble: 10 × 2
##       x       y
##   <dbl> <dbl>
## 1  55.4  97.2
## 2  51.5  96.0
## 3  46.2  94.5
## 4  42.8  91.4
## 5  40.8  88.3
## 6  38.7  84.9
## 7  35.6  79.9
## 8  33.1  77.6
## 9  29.0  74.5
## 10 26.2  71.4
```

```
mean(my_data$x)
```

```
## [1] 54.26327
```

```
mean(my_data$y)
```

```
## [1] 47.83225
```

```
cor(my_data$x, my_data$y)
```

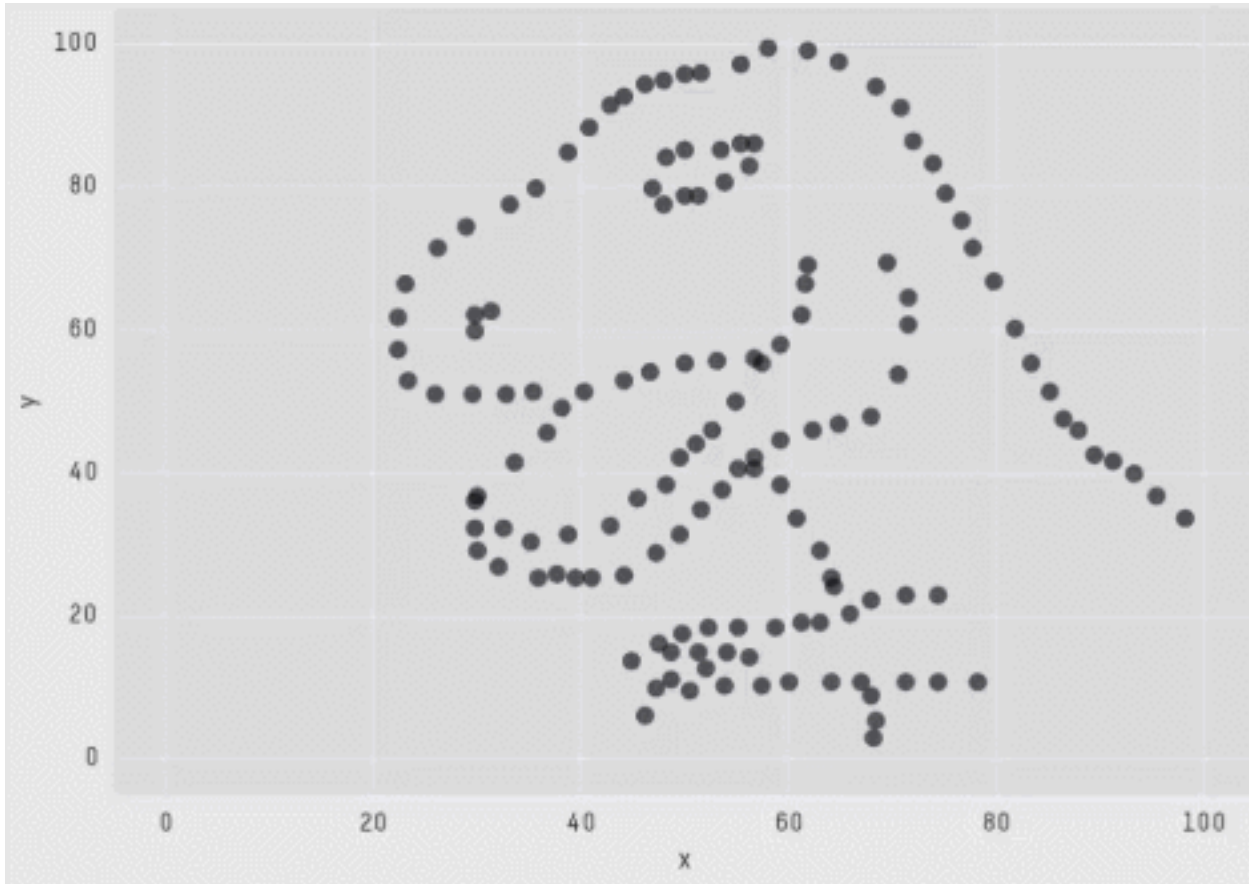
```
## [1] -0.06447185
```

Seems reasonable

Seems reasonable

No correlation

oh no

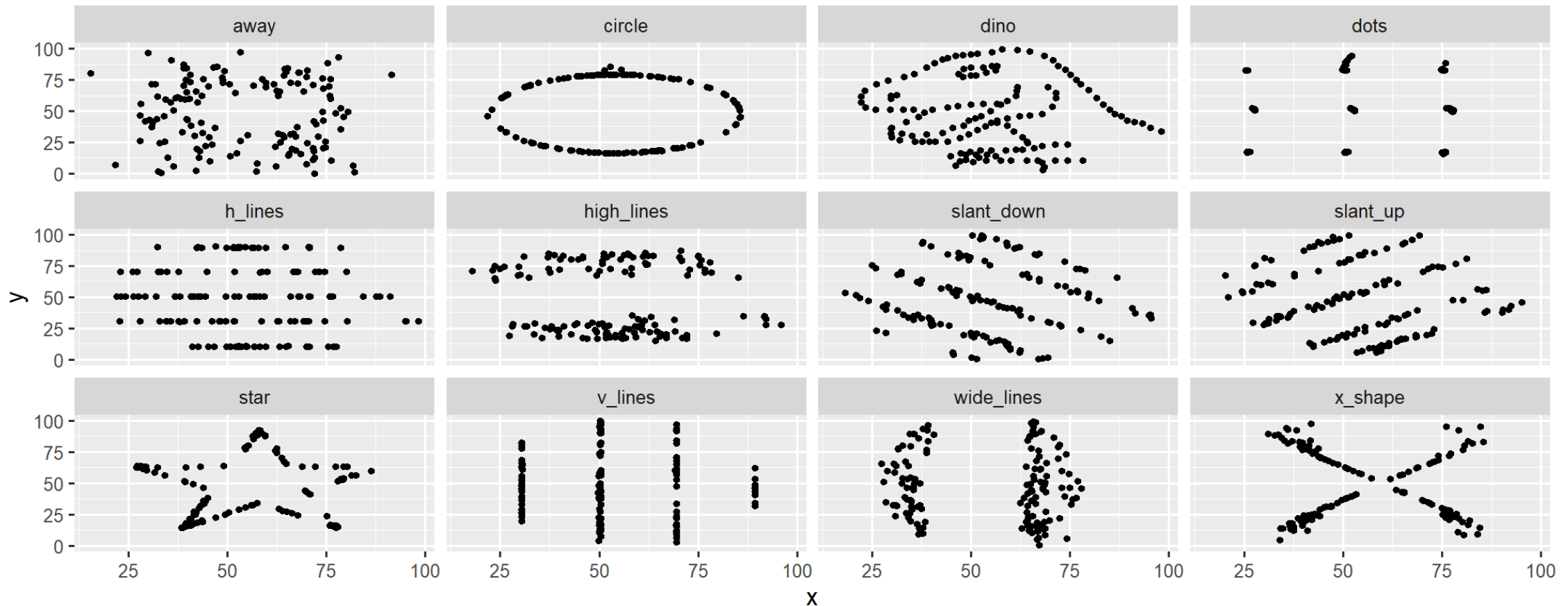


X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526

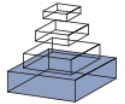
The Datasaurus Dozen

Raw data is not enough

Each of these has the same mean, standard deviation, variance, and correlation



Humans love patterns



Superior pattern processing is the essence of the evolved human brain

Mark P. Mattson^{1,2*}

¹ Laboratory of Neurosciences, National Institute on Aging Intramural Research Program, Baltimore, MD, USA

² Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Edited by:

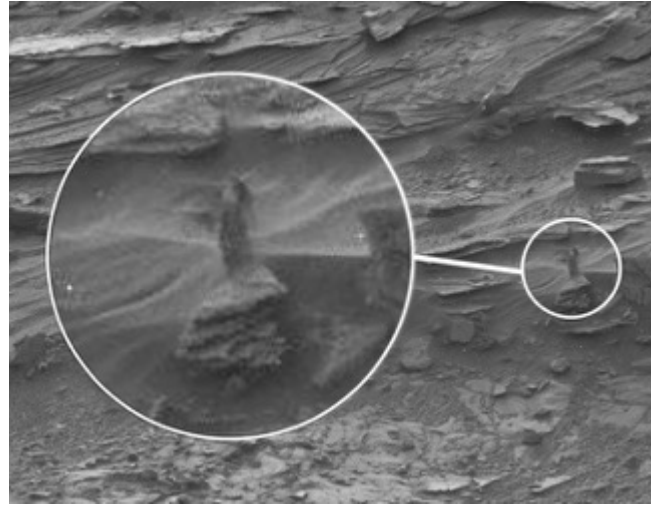
J. Michael Williams, Drexel
University, USA

Humans have long pondered the nature of their mind/brain and, particularly why its capacities for reasoning, communication and abstract thought are far superior to other species, including closely related anthropoids. This article considers superior pattern

<https://doi.org/10.3389/fnins.2014.00265>

(Sometimes we love them too much)

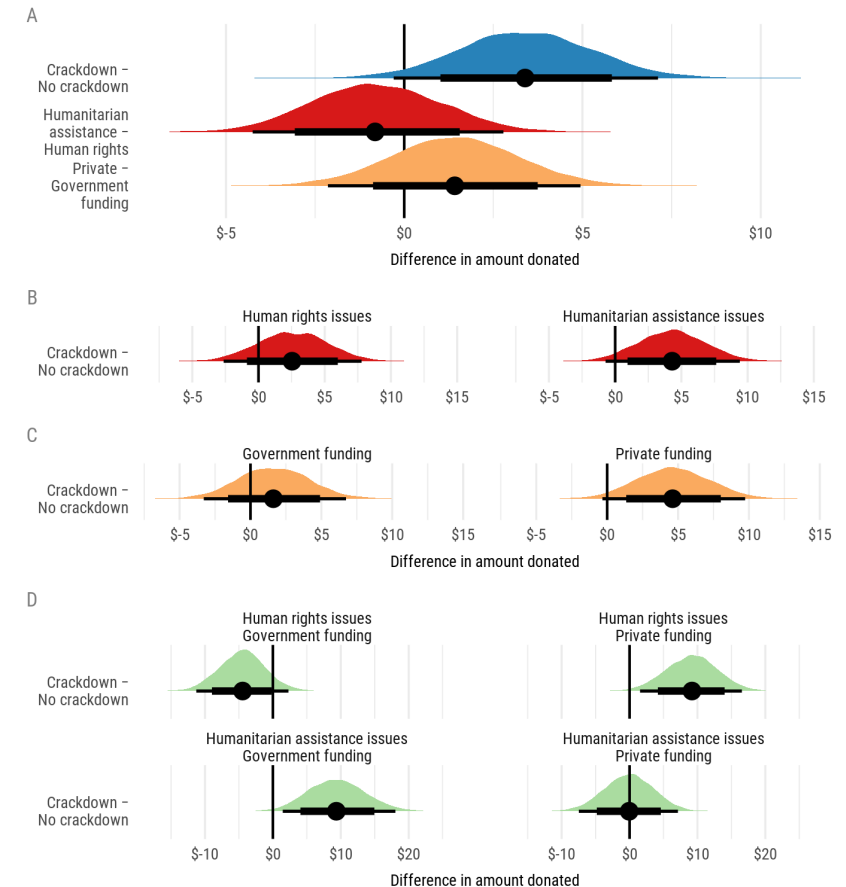
Pareidolia: seeing patterns that aren't there.



Beauty is necessary to see patterns

Table 2: Mean values and differences in means for amount donated in “crackdown” (treatment) and “no crackdown” (control) conditions; values represent posterior medians

H _{1b}	Amount _{Treatment}	Amount _{Control}	Δ	% Δ	$p(\Delta \neq 0)$
Crackdown – No crackdown	16.34	12.93	3.39	26.3%	0.97
Humanitarian assistance – Human rights	14.06	14.85	-0.82	-5.5%	0.67
Private – Government funding	15.13	13.71	1.42	10.4%	0.79
H _{2b} and H _{3b}	Amount _{Crackdown}	Amount _{No crackdown}	Δ	% Δ	$p(\Delta \neq 0)$
Human rights issues	17.4	14.86	2.54	17.2%	0.83
Humanitarian assistance issues	15.91	11.68	4.3	36.9%	0.95
Government funding	13.83	12.24	1.61	13.1%	0.74
Private funding	18.95	14.23	4.62	32.4%	0.97
H _{2b} and H _{3b} (nested)	Amount _{Crackdown}	Amount _{No crackdown}	Δ	% Δ	$p(\Delta \neq 0)$
Human rights issues, Government funding	10.56	15.15	-4.46	-29.5%	0.91
Human rights issues, Private funding	23.76	14.5	9.19	63.8%	0.99
Humanitarian assistance issues, Government funding	21.42	11.89	9.35	77.9%	0.99
Humanitarian assistance issues, Private funding	15.69	15.72	-0.05	-0.3%	0.51



Point shows posterior median; thick black lines show 80% credible interval; thin black lines show 95% credible interval

Beautiful visualizations

What makes a great visualization?

Truthful

Functional

Beautiful

Insightful

Enlightening

Alberto Cairo, *The Truthful Art*

What makes a great visualization?

"Graphical excellence is the well-designed presentation of interesting data—a matter of substance, of statistics, and of design ... [It] consists of complex ideas communicated with clarity, precision, and efficiency. ... [It] is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space ... [It] is nearly always multivariate ... And graphical excellence requires telling the truth about the data."

Edward Tufte, *The Visual Display of Quantitative Information*, p. 51

What makes a great visualization?

Good aesthetics

No substantive issues

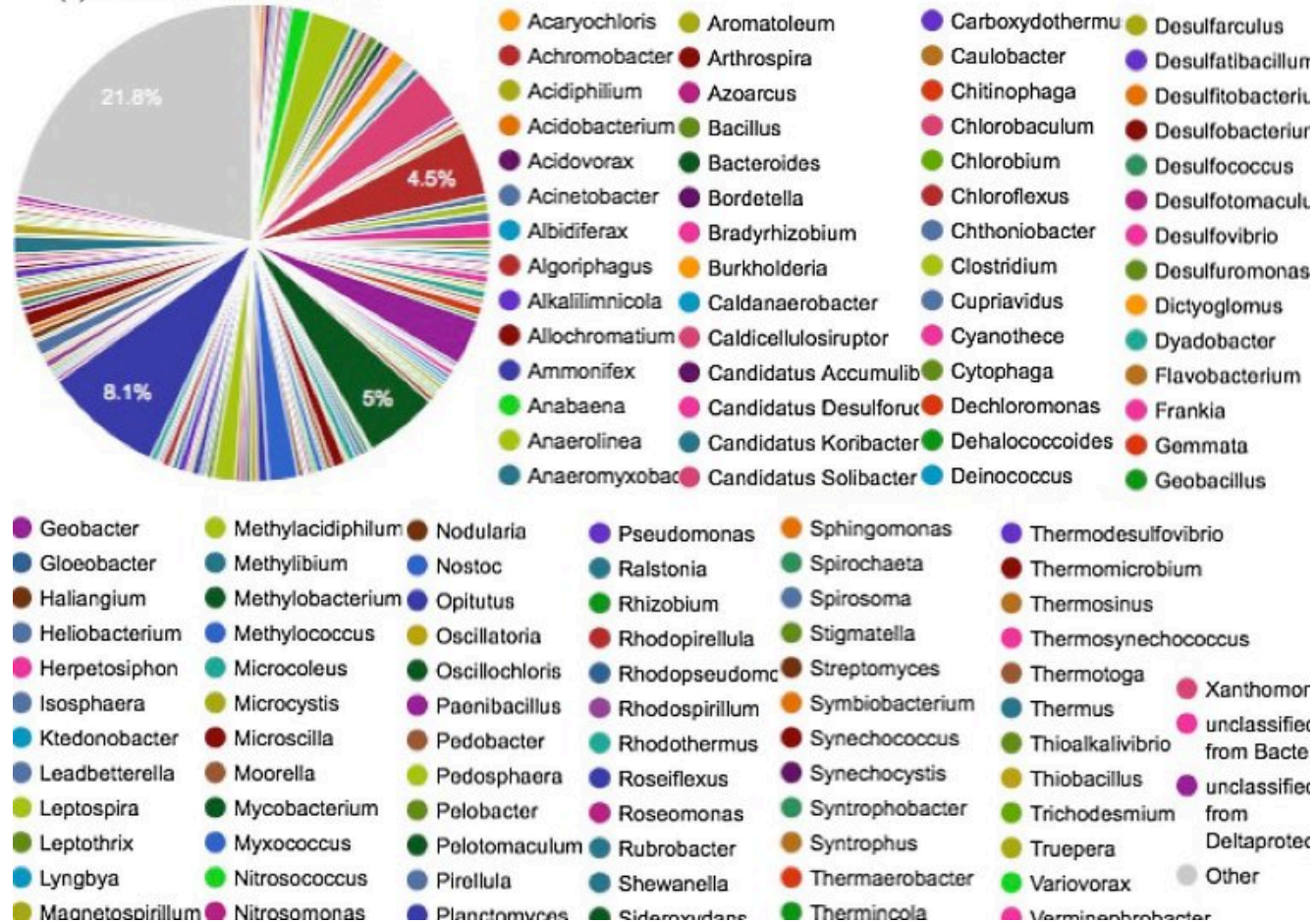
No perceptual issues

Honesty + good judgment

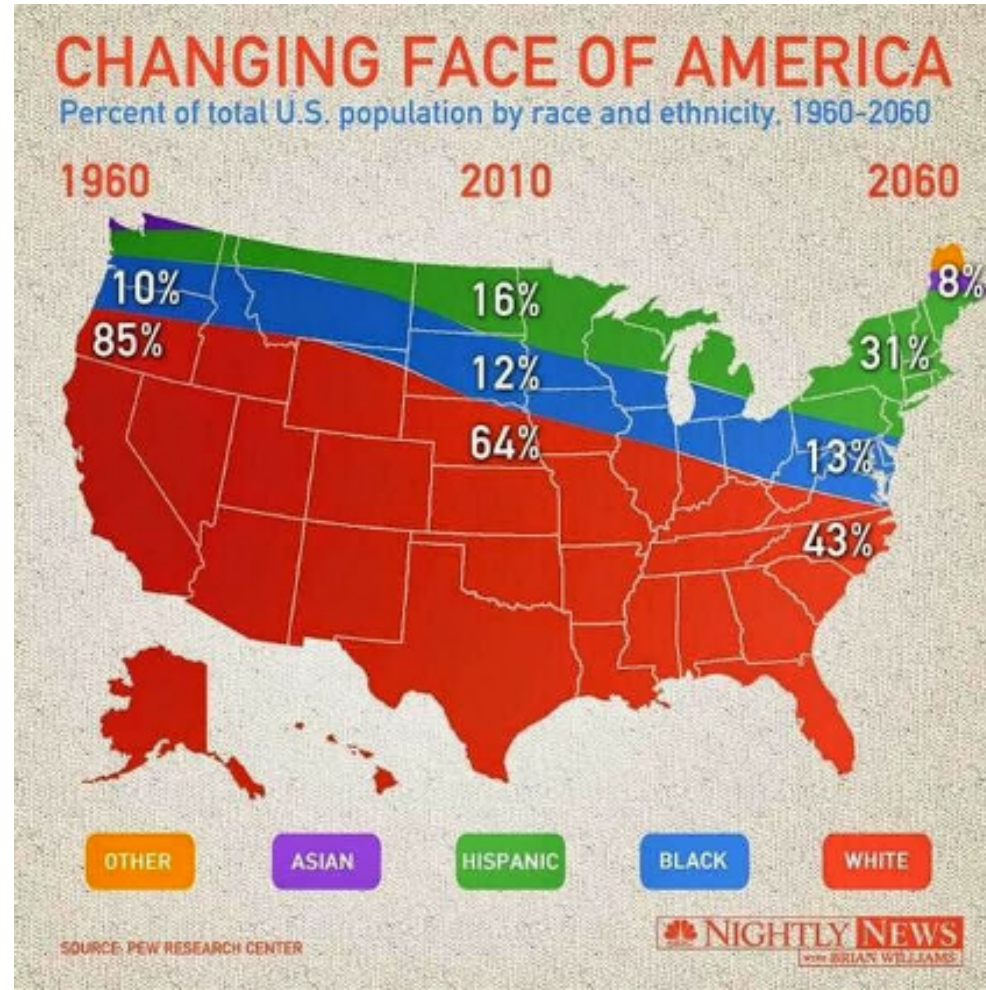
Kieran Healy, Data Visualization: A Practical Introduction

What's wrong?

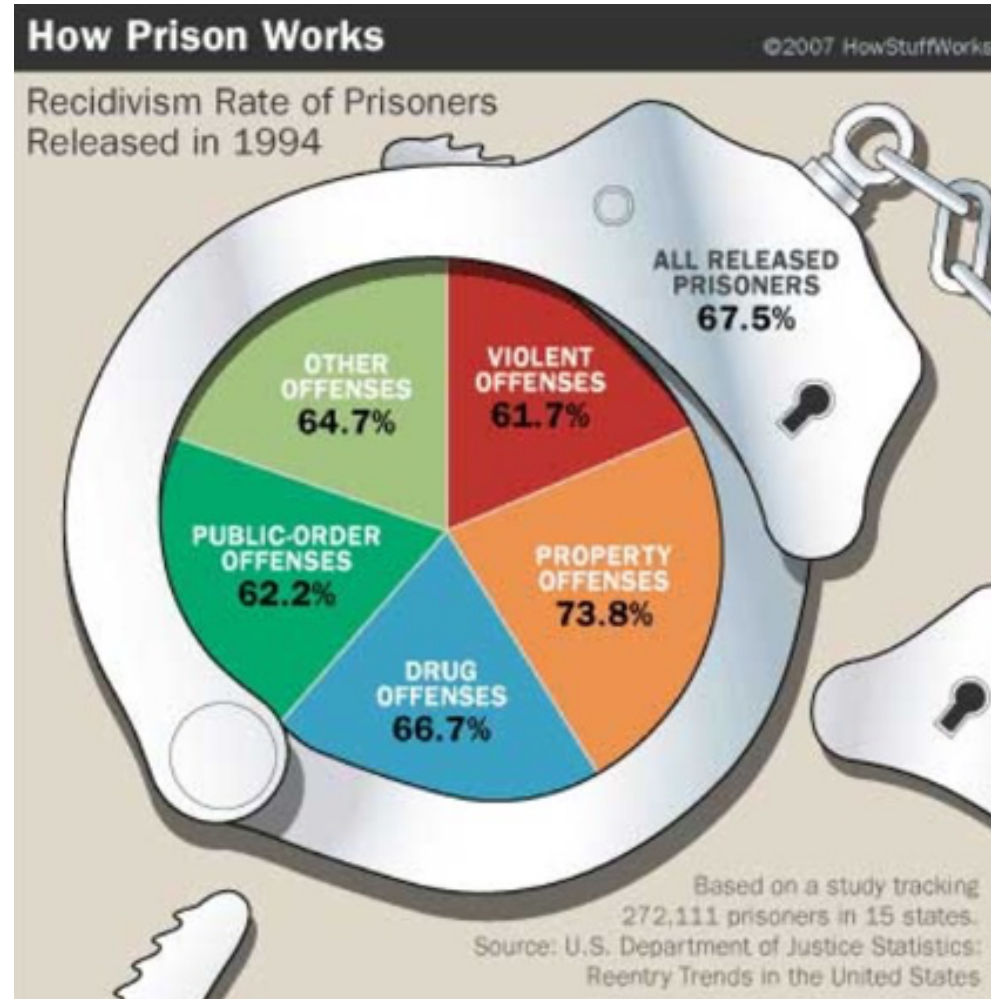
(f) Distribution of Genus



What's wrong?



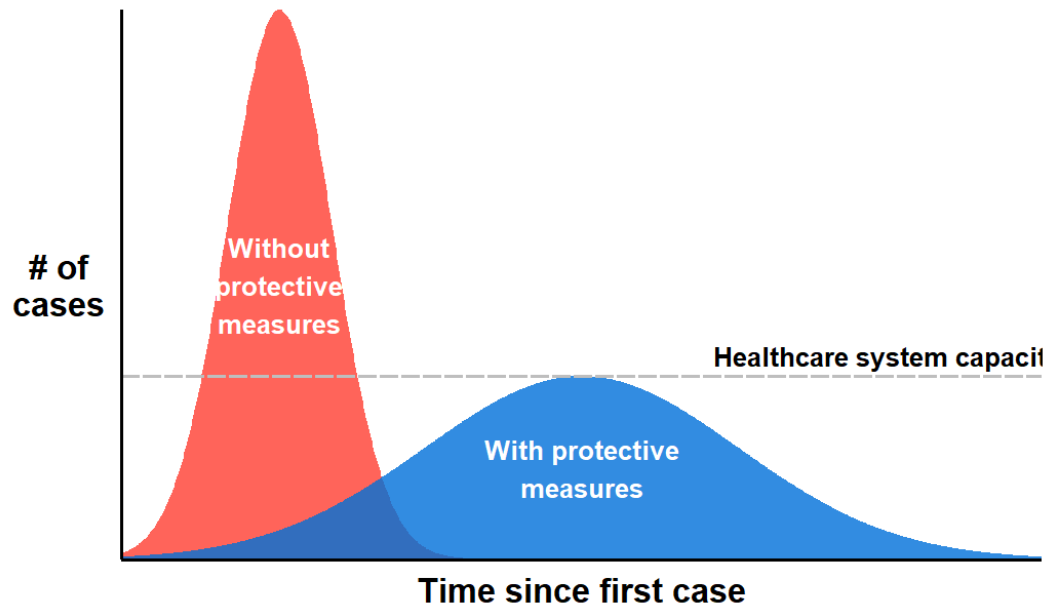
What's wrong?



What's right?

Flatten the curve!

Slow down community spread by social distancing



Adapted from the CDC and The Economist
Visit flattenthecurve.com

Carl T. Bergstrom @CT_Bergstrom · Mar 6

3. There is a lot of complicated epidemiological modeling behind this idea, but this graphic strips all of that away, and discards irrelevant details to provide a straightforward story that people find easy to grasp at a glance.

It **simplifies** and **highlights** what matters.

6 198 1.8K

[Show replies](#)

Carl T. Bergstrom @CT_Bergstrom · Mar 6

4. I've seldom seen a piece of sci-comm matter so much. We have an opportunity to flatten the [#COVID19](#) [#coronavirus](#) epidemic curve by aggressive social distancing and other measures.

But people don't understand what the point is, if the virus is going to circulate broadly.

8 313 2K

[Show replies](#)

Carl T. Bergstrom @CT_Bergstrom · Mar 6

5. This graph provides the answer, powerfully and concisely.

And because of that, it has exploded across twitter and other media. I've used it myself a number of times. This graph is changing minds, and by changing minds, it is saving lives.

6 196 1.5K

Thread by Carl T. Bergstrom