

Seeing is Conceiving

Brian A. Fannin

26 September 2019

Overview

Where to find this

This presentation may be found at:

http://pirategrunt.com/giro_2019

Code to produce the examples and slides:

https://github.com/PirateGrunt/giro_2019

What we'll talk about

- ▶ Semiotics
- ▶ Information density
- ▶ BONUS: What's new in CAS research

Semiotics

Seeing is believing

conceive (v) - form a mental representation of; imagine.

Seeing is *conceiving*

conceive (v) - form a mental representation of; imagine.

For statisticians there always have to be comparisons; numbers on their own are not enough.

- ▶ Gelman and Unwin

Which of these two numbers is larger?

11

9

How about these two?

1011

1001

These?

IX

XI

And these?

9

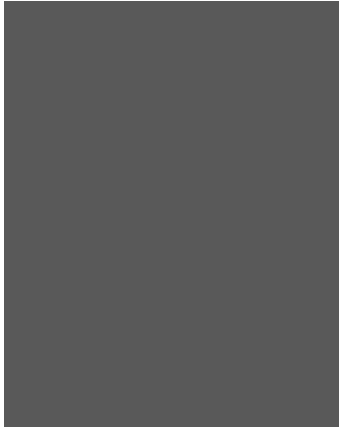
B

These?

11

9

How about these?



nine

9

neun

1001

IX

nueve

Arabic or sanskrit are no more legitimate than any other representation of numbers.

Be prepared to accept the idea that there are circumstances when geometric primitives may be understood *faster*.

Another comparison

999999999

99999999999

These two?

999,999,999

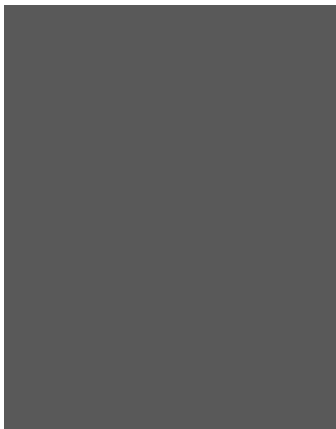
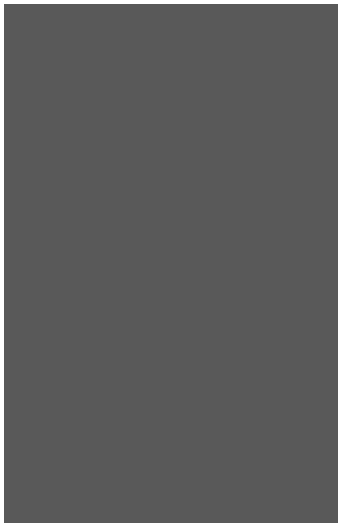
99,999,999,999

What does an order of magnitude look like?



Information density

This is actually *too much* information



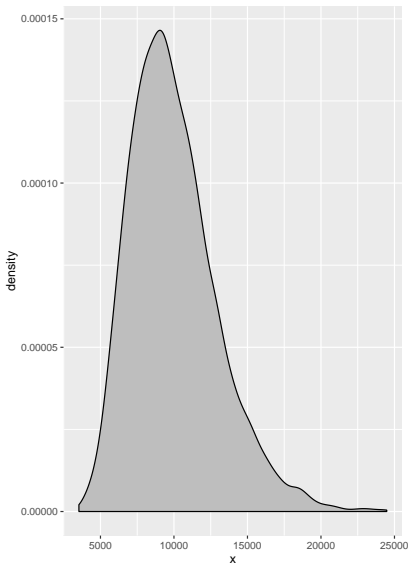
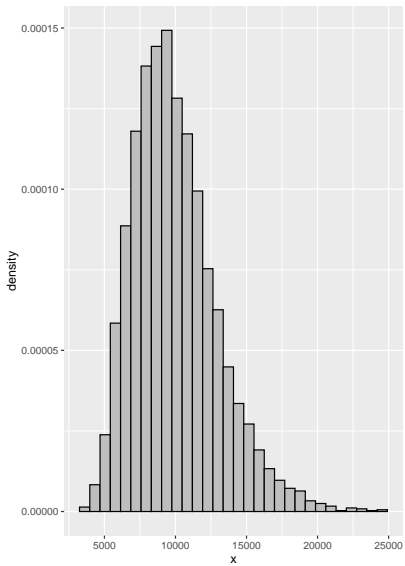
This is better

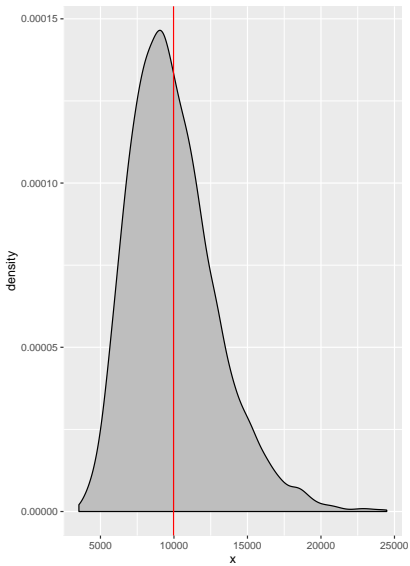
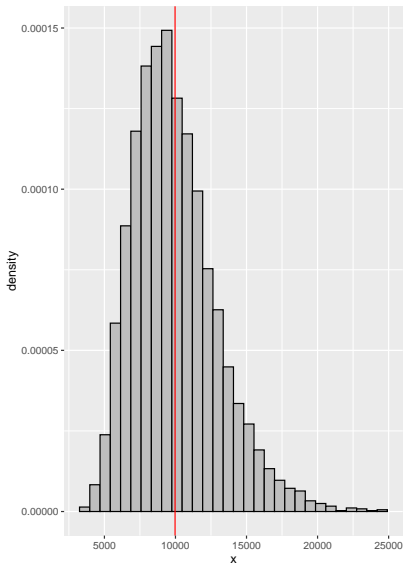


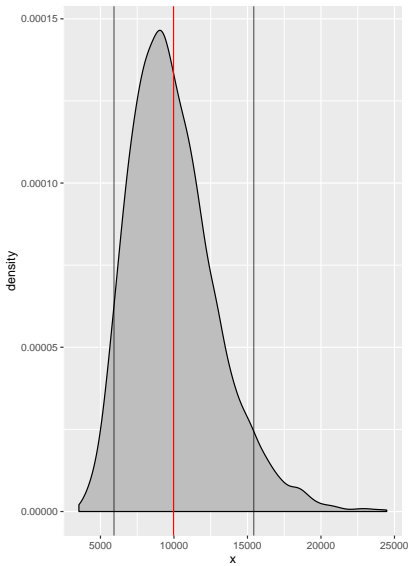
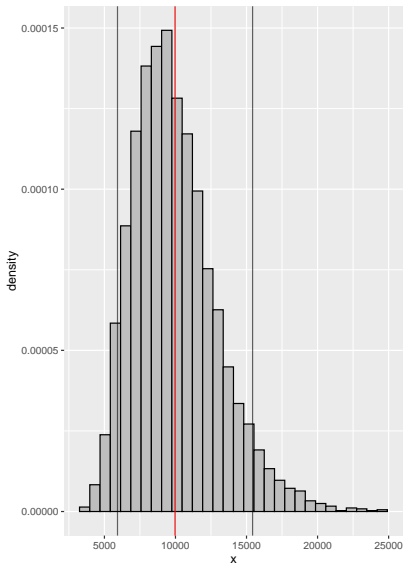
Statistics maps a set of many numbers into a set of fewer numbers.

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	3534	7854	9562	9973	11644	24476

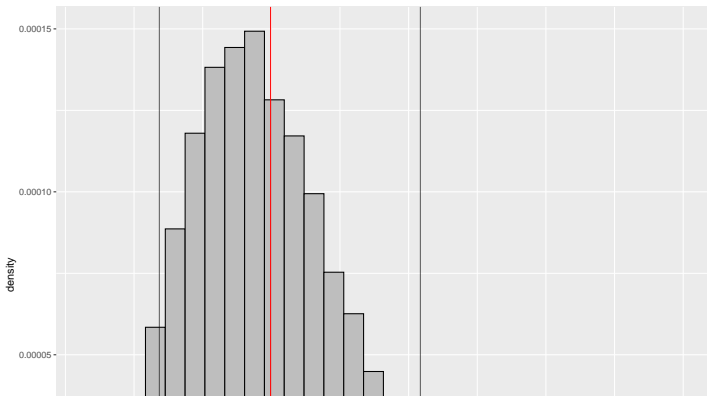
metric	x
Min.	3,534
1st Qu.	7,854
Median	9,562
Mean	9,973
3rd Qu.	11,644
Max.	24,476







metric	x
Min.	3,534
1st Qu.	7,854
Median	9,562
Mean	9,973
3rd Qu.	11,644
Max.	24,476

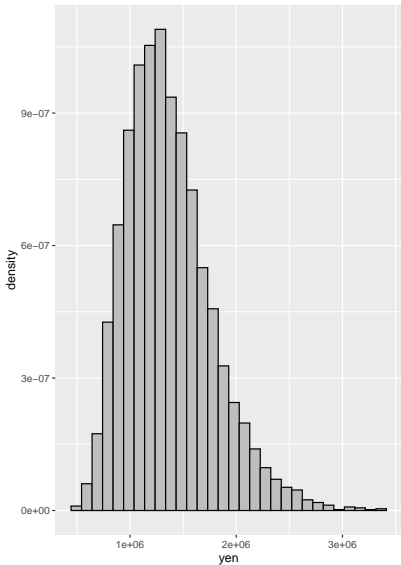
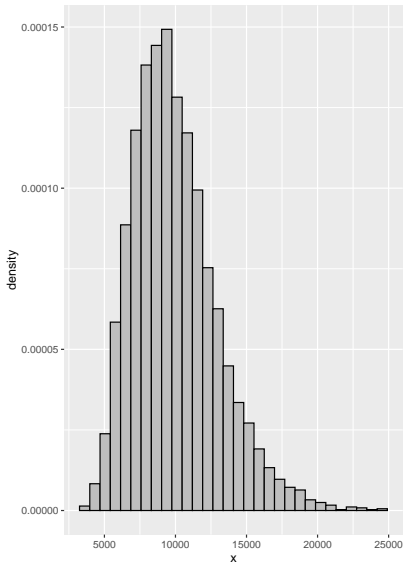


Summary statistics are *always* a reduction of information.

Visualization presents (almost) *all* of the data. The reductions are made with our eyes.

Which data set has greater skew?

metric	x	z
Min.	3,534	484,214
1st Qu.	7,854	1,076,044
Median	9,562	1,309,932
Mean	9,973	1,366,328
3rd Qu.	11,644	1,595,286
Max.	24,476	3,353,189

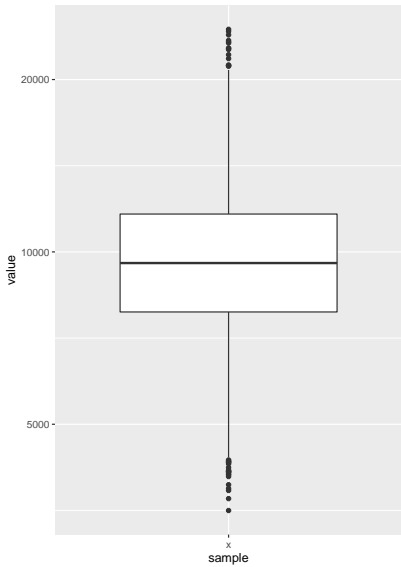
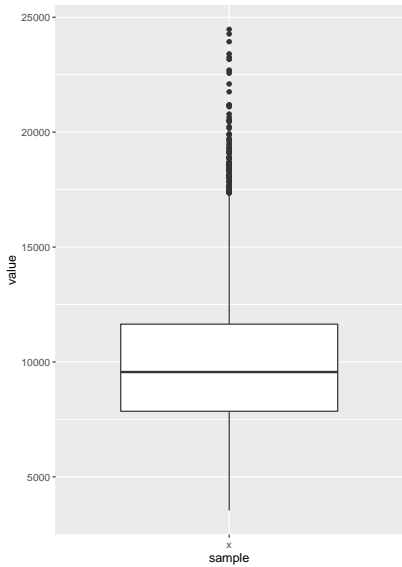


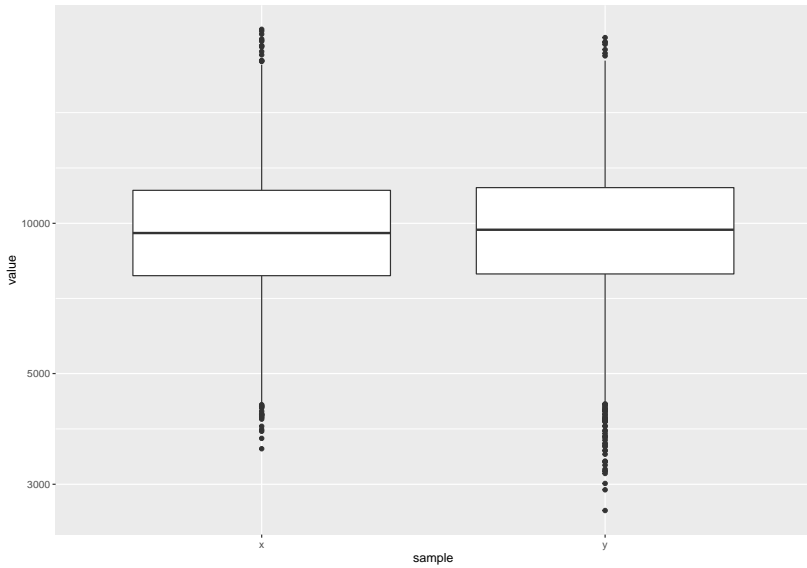
metric	x	z
Min.	3,534	484,214
1st Qu.	7,854	1,076,044
Median	9,562	1,309,932
Mean	9,973	1,366,328
3rd Qu.	11,644	1,595,286
Max.	24,476	3,353,189

metric	x	z
skew	0.8532277	0.8532277

Visualization as a metaphor for a model

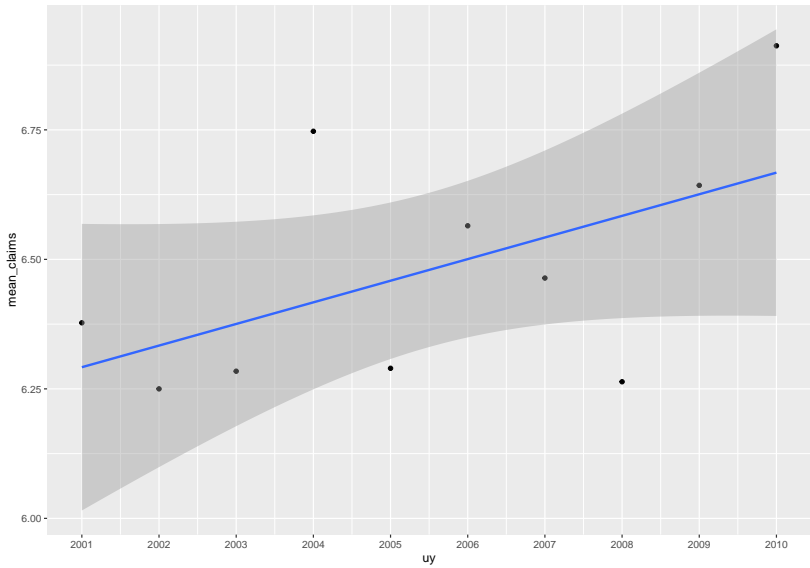
metric	x	y
Min.	3,534	2,658
1st Qu.	7,854	7,917
Median	9,562	9,709
Mean	9,973	9,986
3rd Qu.	11,644	11,791
Max.	24,476	23,562

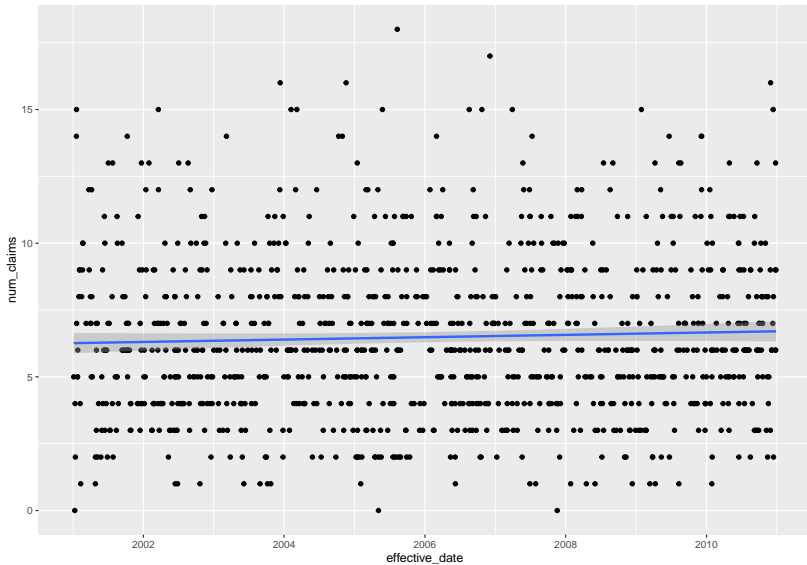




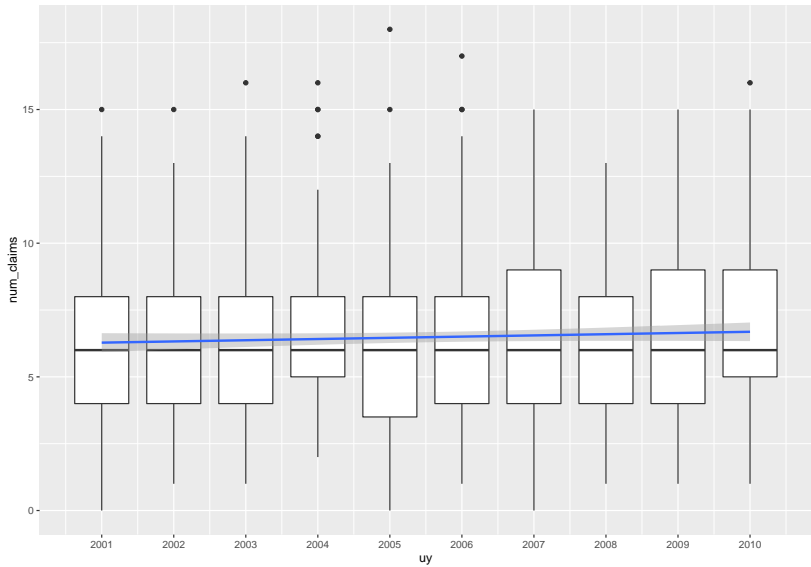
Financial reporting view

uy	mean_claims
2001	6.4
2002	6.2
2003	6.3
2004	6.7
2005	6.3
2006	6.6
2007	6.5
2008	6.3
2009	6.6
2010	6.9

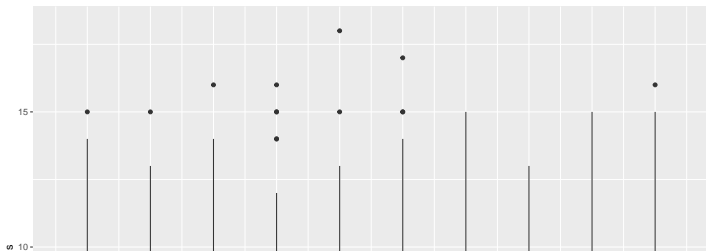




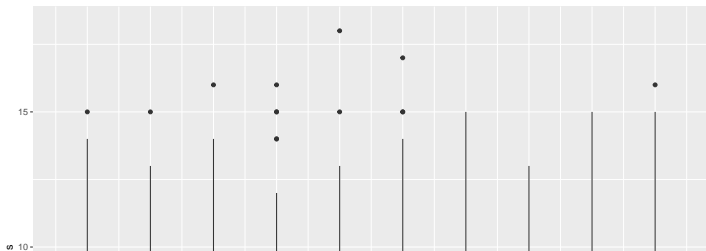
effective_date	num_claims
2002-08-26	4
2008-02-25	4
2005-08-09	6
2007-05-21	6
2005-07-23	2
2010-06-22	7
2009-11-01	4
2008-10-10	5
2001-06-28	2
2007-05-24	13
2002-06-24	3
2005-03-29	6
2001-05-30	3
2004-11-14	5
2007-12-18	10

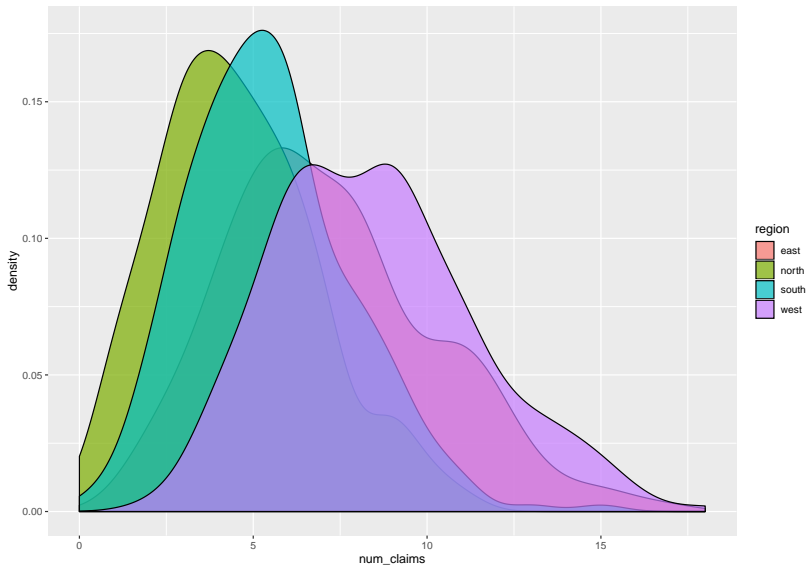


uy	mean_claims	median_claims
2001	6.4	6
2002	6.2	6
2003	6.3	6
2004	6.7	6
2005	6.3	6
2006	6.6	6
2007	6.5	6
2008	6.3	6
2009	6.6	6
2010	6.9	6



uy	mean_claims	median_claims
2001	6.38	6
2002	6.25	6
2003	6.28	6
2004	6.75	6
2005	6.29	6
2006	6.56	6
2007	6.46	6
2008	6.26	6
2009	6.64	6
2010	6.91	6





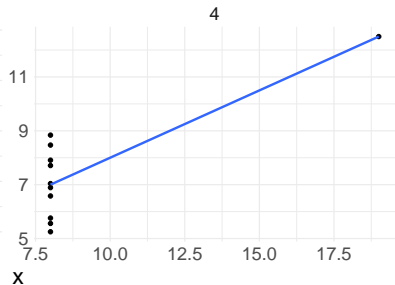
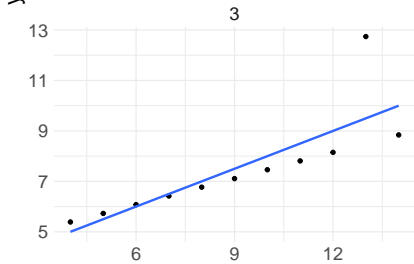
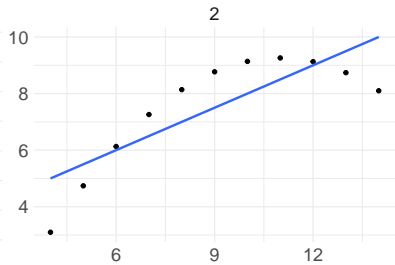
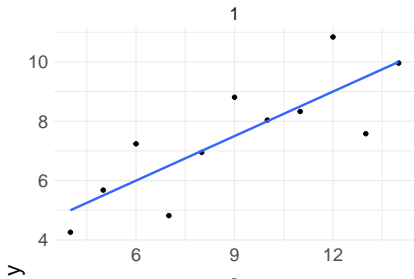
Validating models

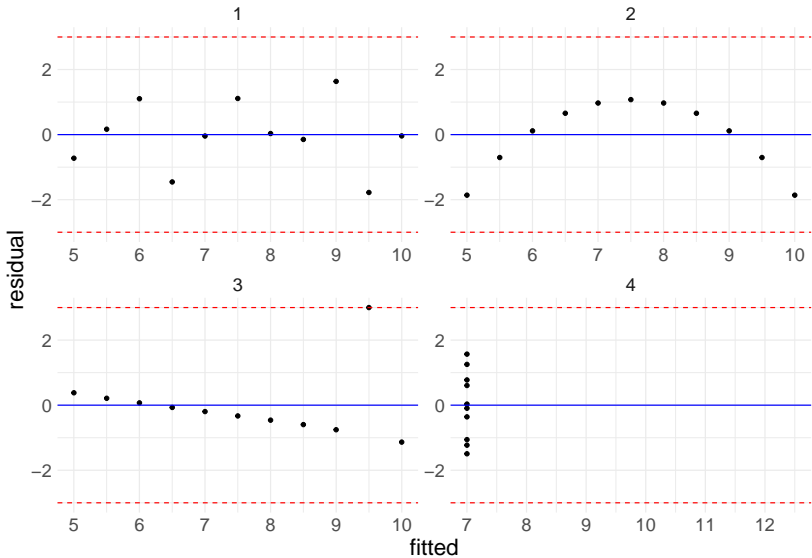
A good model is one which isolates signal and leaves only noise.

->

Are these good models?

segment	adj.r.squared	sigma
1	62.9%	1.2
2	62.9%	1.2
3	62.9%	1.2
4	63.0%	1.2

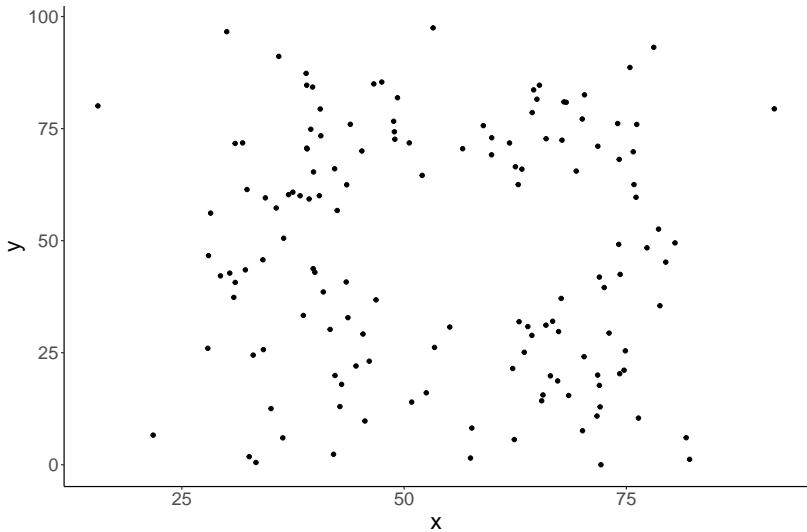




Correlation is not causation

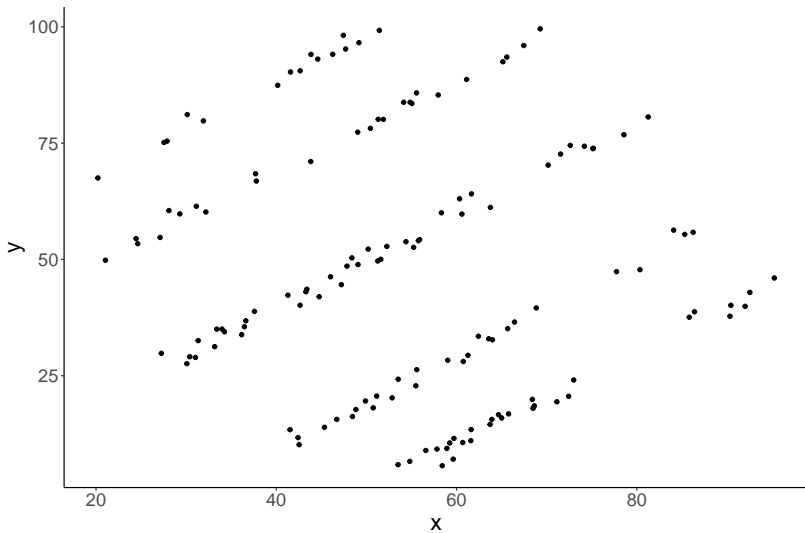
set	correlation
a	-6.4%
b	-6.9%
c	-6.8%
d	-6.4%
e	-6.0%
f	-6.2%
g	-6.9%
h	-6.9%
i	-6.9%
j	-6.3%
k	-6.9%
l	-6.7%
m	-6.6%

Correlation = -6.4%



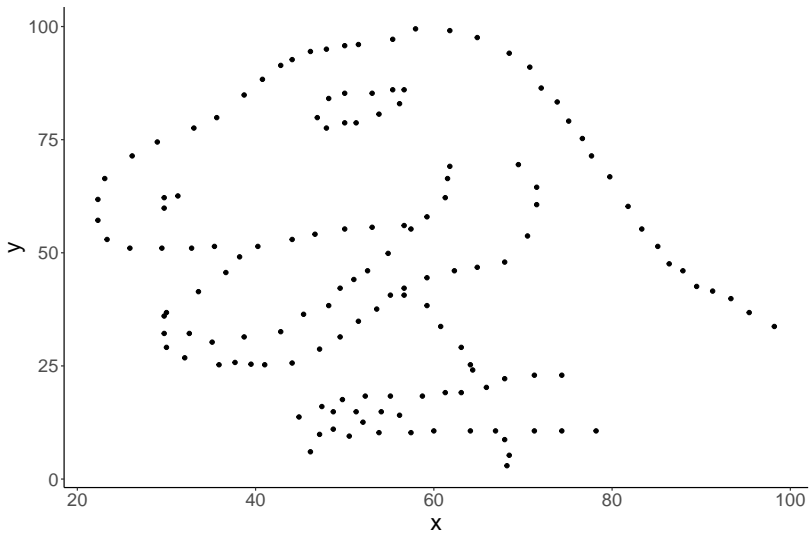
Another

Correlation = -6.9%



Another

Correlation = -6.4%



Seeing is conceiving

What's new with CAS research?

Wrapping up

- ▶ Exposure bases for cyber liability
- ▶ Compartmental reserving methods
- ▶ Machine learning in reserving

On deck

- ▶ Economic scenario generators for non-life insurance
- ▶ Flood modeling using public data
- ▶ On the viability of microinsurance in North America

New GitHub account

<https://github.com/casact>

The screenshot shows a web browser window with the URL `github.com/casact`. The page header includes the organization name `casact` and navigation links for `Repositories`, `Packages`, `People`, `Teams`, `Projects`, and `Settings`. The profile section features the `CAS` logo, the name `Casualty Actuarial Society`, and a bio: "This will serve as a platform for development of tools that prove useful to non-life actuaries." Below the bio are links for the website (`https://www.casact.org/`) and email (`github@casact.org`).

The `Pinned repositories` section displays a list of repositories. The first pinned repository is `raw_package`, described as "Data package for R actuarial workshops". It has 2 stars and 1 fork. Below the list is a search bar labeled "Find a repository..." and filters for "Type: All" and "Language: All". A green `New` button is also present.

Below the search bar, the `raw_package` repository is shown in detail, including its description, 1 star, 0 forks, and 0 watchers, and an update date of "Apr 26, 2018". Below this, the `OSSCIntro` repository is shown, described as "A document that will be used to collaborate on learning collaborative editing using repositories", with 1 star, 0 forks, and 0 watchers, updated on "Dec 11, 2013".

On the right side of the page, there are two sections: `Top languages`, showing `R` and `TeX` as the top languages, and `People`, showing 2 people and an `Invite someone` button.

Talk to me

bfannin@casact.org

Thank you!

Where to find this

This presentation may be found at:

http://pirategrunt.com/giro_2019

Code to produce the examples and slides:

https://github.com/PirateGrunt/giro_2019