Data Visualizations

Marco Morales
marco.morales@columbia.edu

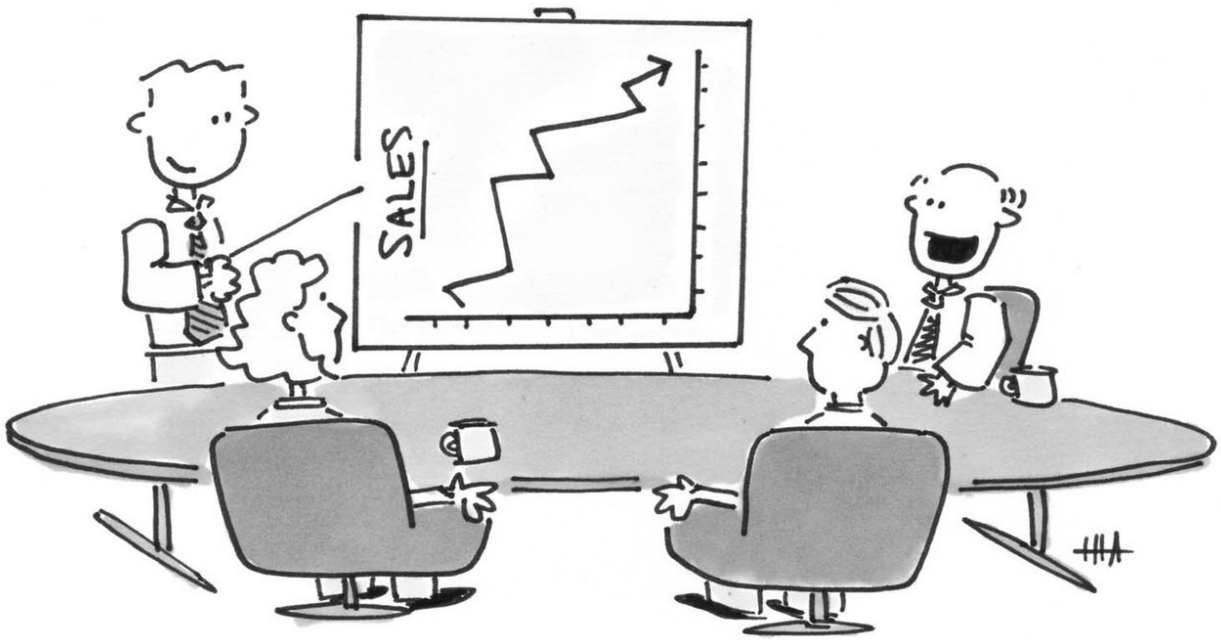
Nana Yaw Essuman
nanayawce@gmail.com

GR5069
Topics in Applied Data Science
for Social Scientists

Spring 2021
Columbia University

*The **power of a graph** is its ability to enable one to take in the quantitative information, organize it, and see patterns and structure **not readily revealed by other means of studying the data**.*

- Cleveland & McGill (1984)



"He's right! When you look at it that way, it's not so bad!"

Data Visualizations

- Also, Gelman is right, data visualizations should be a continuum over
  1. **data exploration**
  2. statistical modeling
  3. **summarizing results / insights**
- **1.** and **3.** require heavier interaction with stakeholders in typical DS world

Data exploration

Visualizations for **data exploration**

- typically, you will work with:
  a) data that is **known to someone**
     - validate your understanding of the data
     - validate correct aggregation / disaggregation
     - leverage someone's dense knowledge
  b) data that is **not known to anyone**
     - understand what is in the data
     - leverage someone else's dense knowledge
- graphs are rarely for yourself, but means to understand / validate data

Visualizations for data exploration

Example: new data on confrontations with organized crime in MX

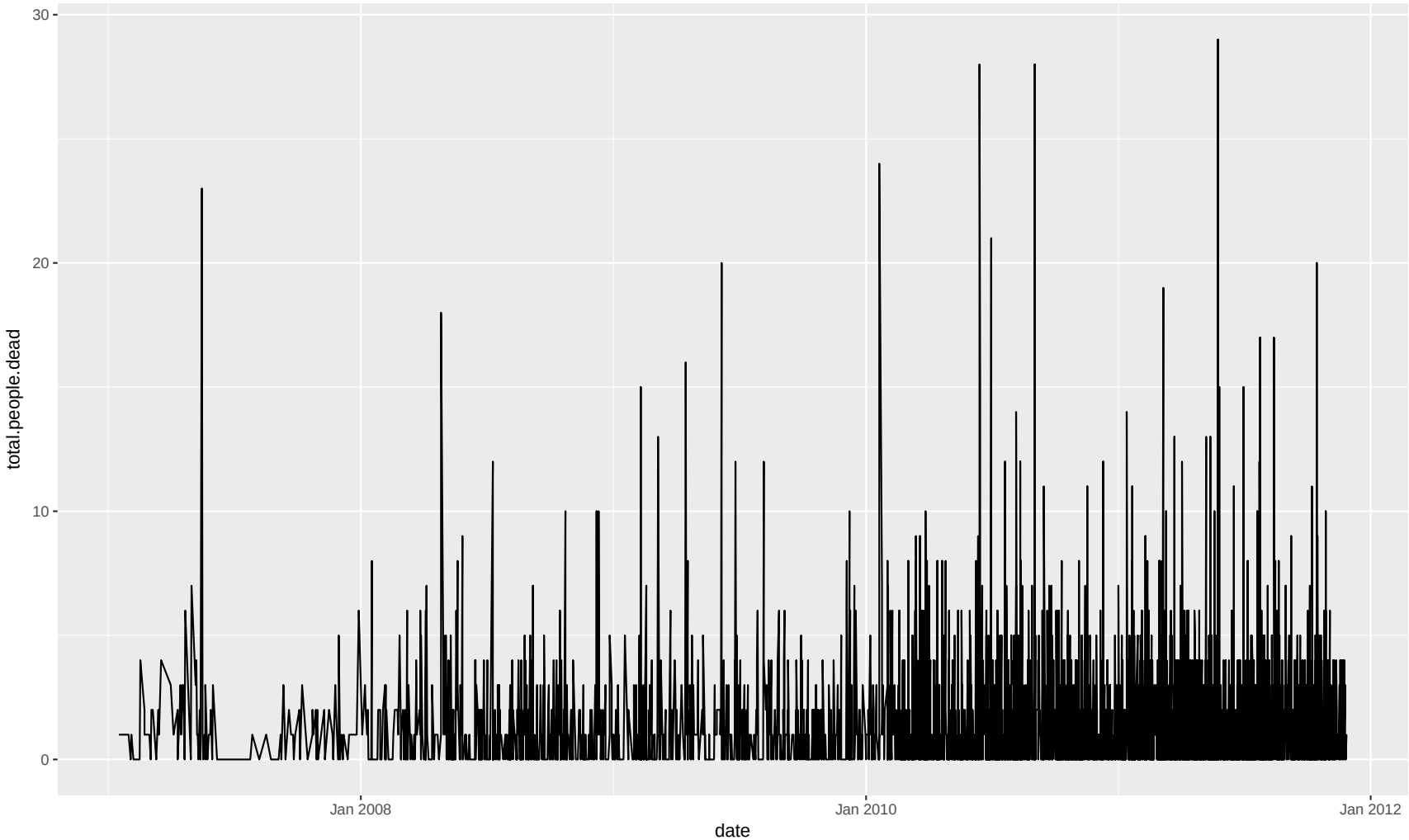
Visualizations for data exploration

Example: new data on confrontations with organized crime in MX

86.1% of dead civilians who presumably participated in confrontations with federal armed forces were killed in events of “perfect lethality" where there were only dead and no wounded. [...] Mexico has the terrible situation of having lethality indices of 2.6. The lethality index of the Federal Police is 2.6 dead for every wounded, the Navy’s reaches 17.3 dead for every wounded, and the Army’s is 9.1 dead for every wounded.

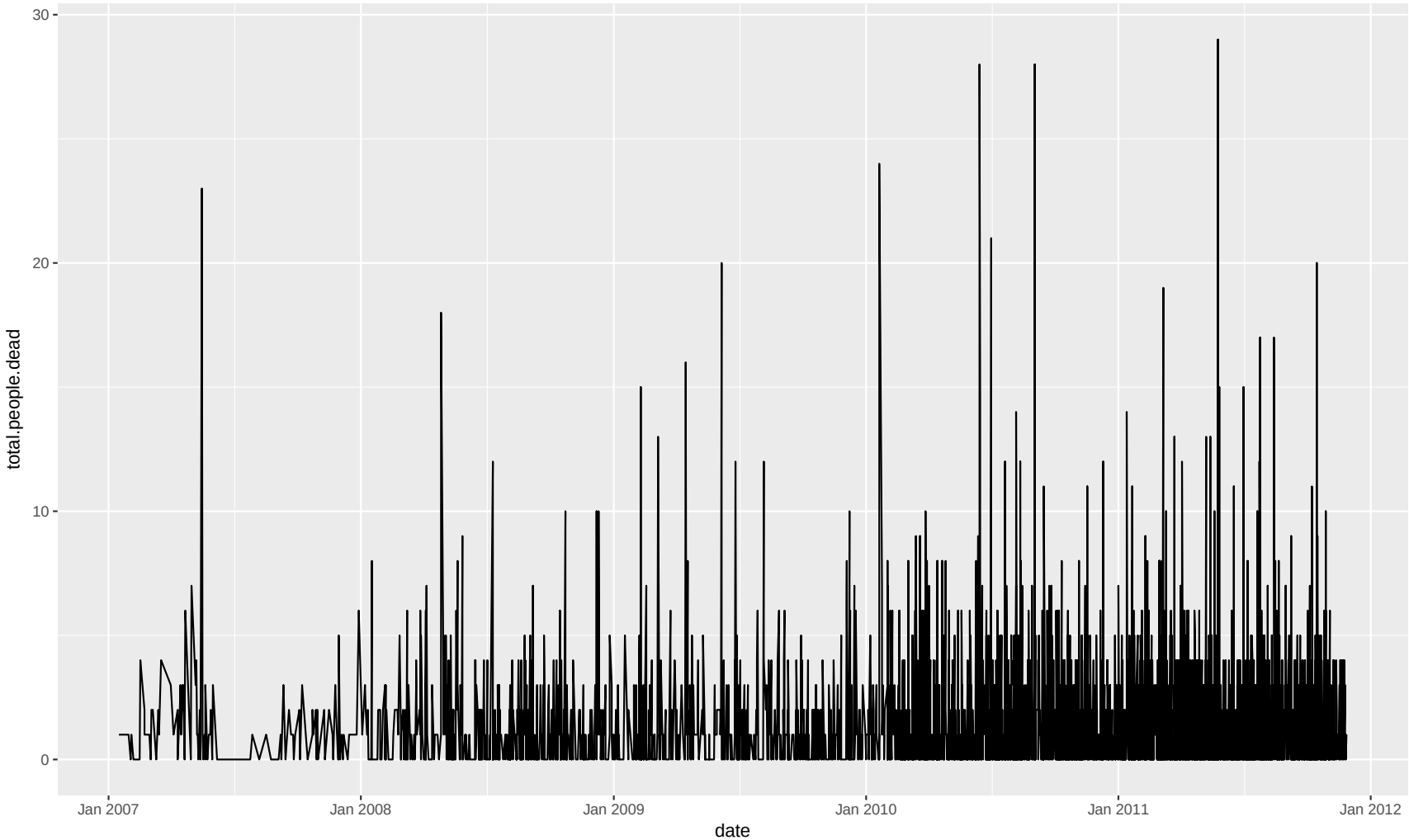
Visualizations for data exploration

let's start simple... it's time-series data after all



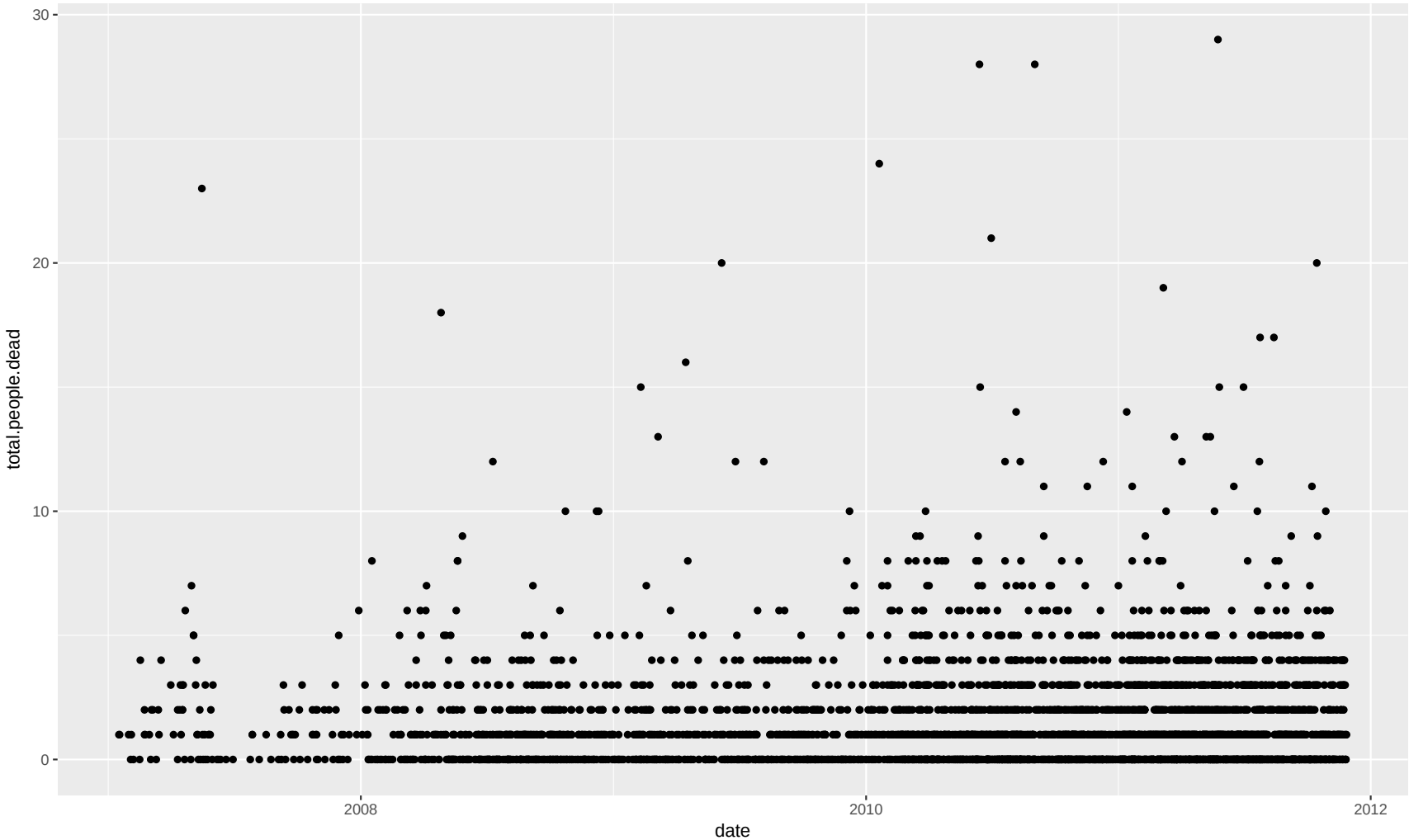
Visualizations for data exploration

if you label it, will it show?



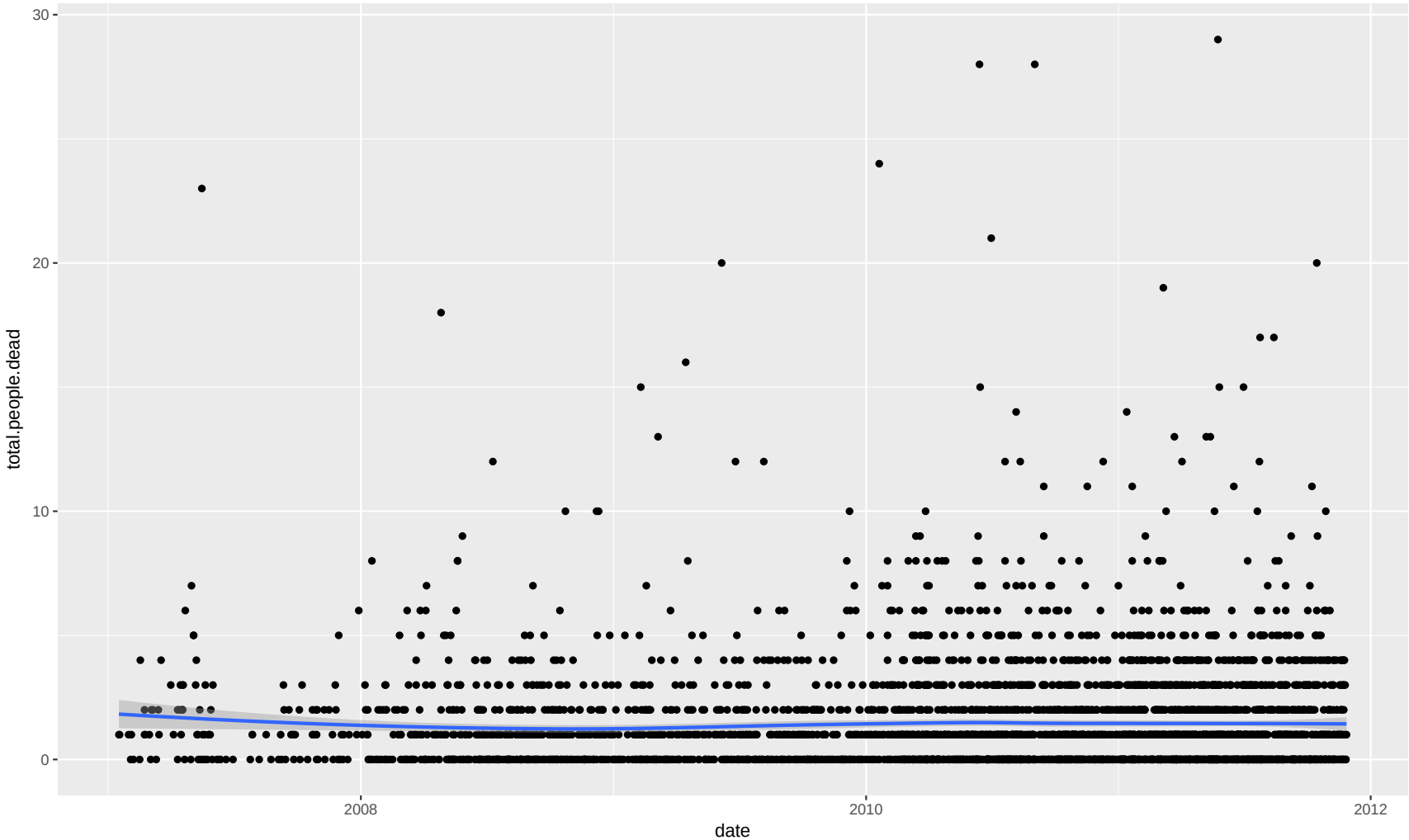
Visualizations for data exploration

perhaps a closer look at the raw data...



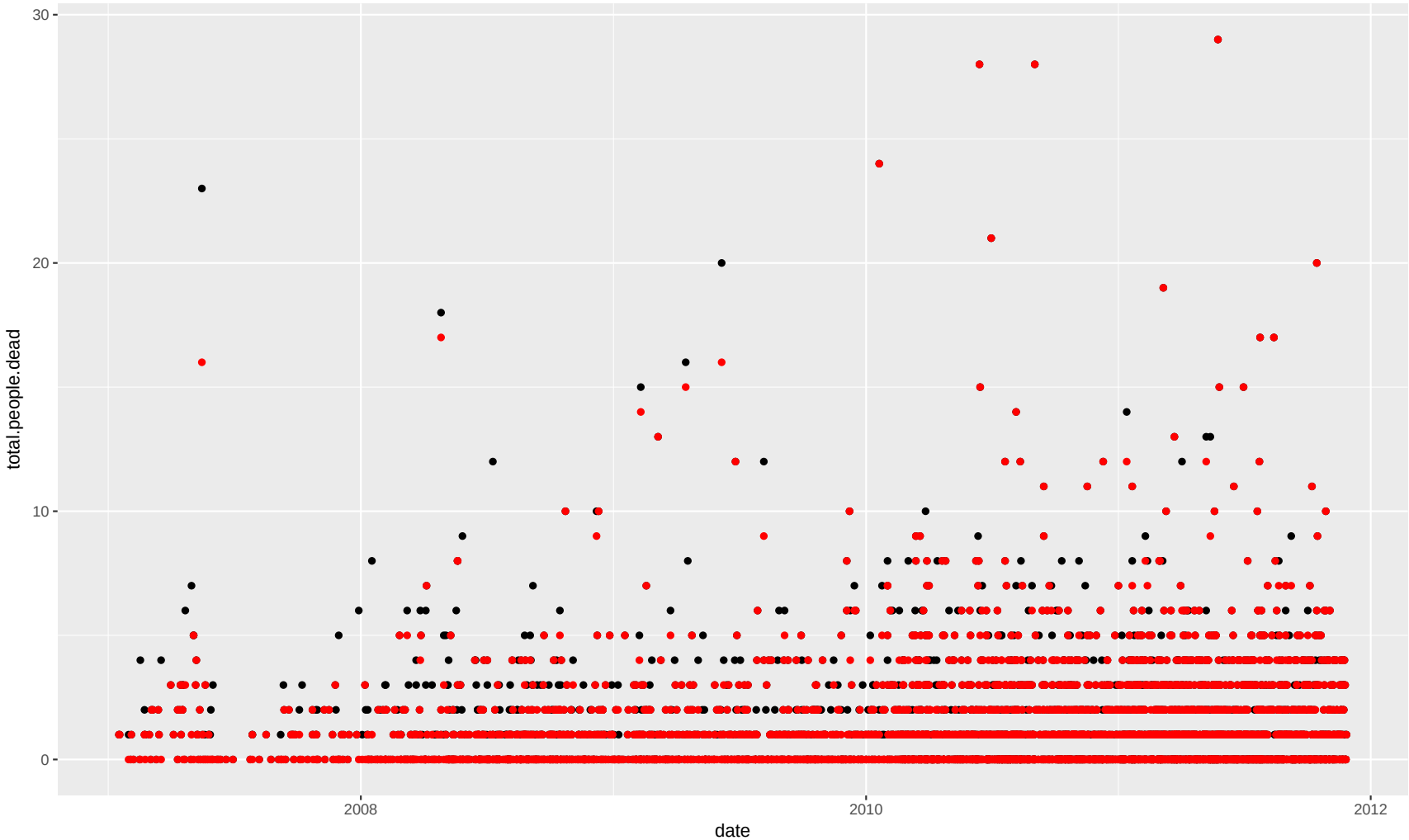
Visualizations for data exploration

too noisy... perhaps a linear pattern?



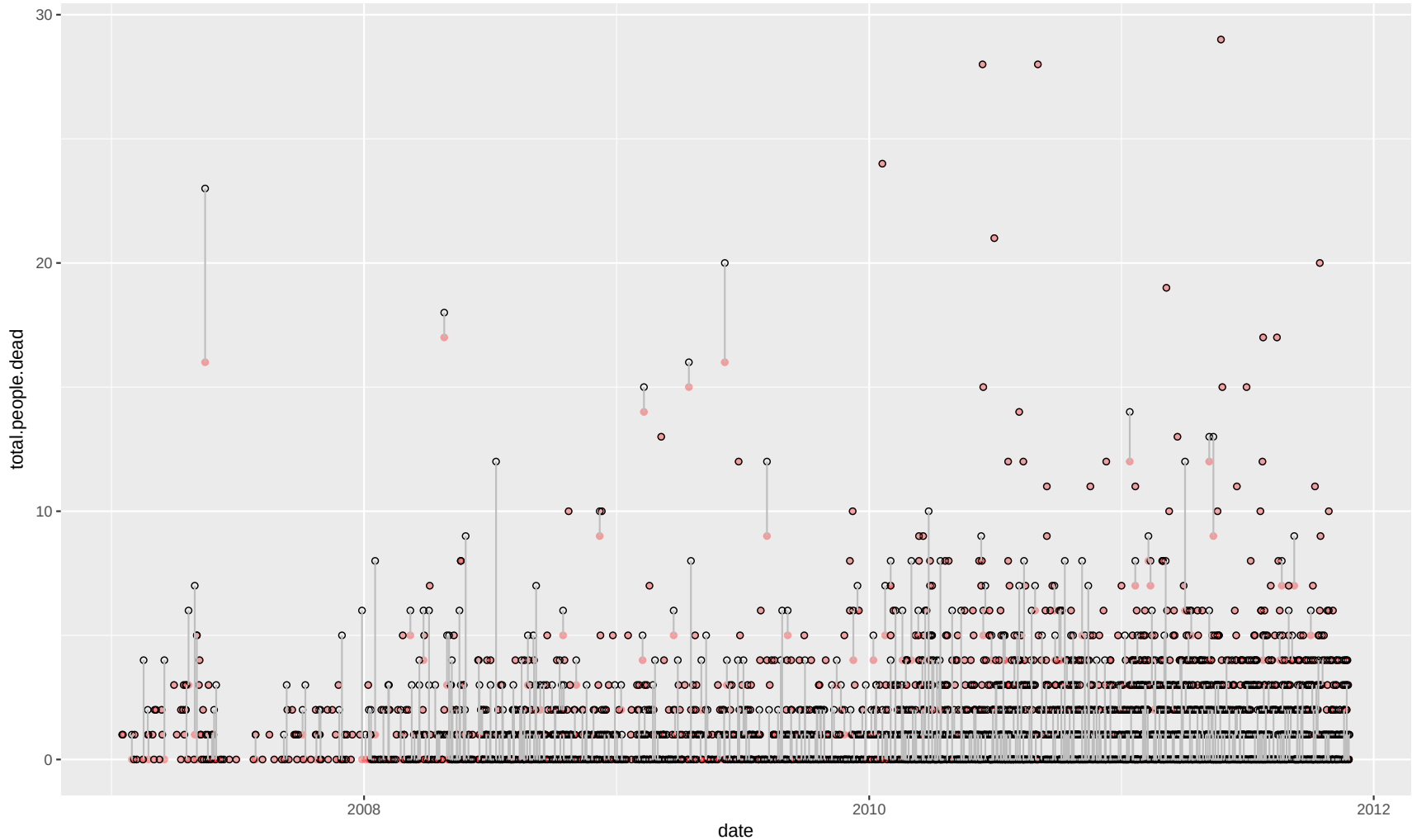
Visualizations for data exploration

let's contrast with organized crime deaths



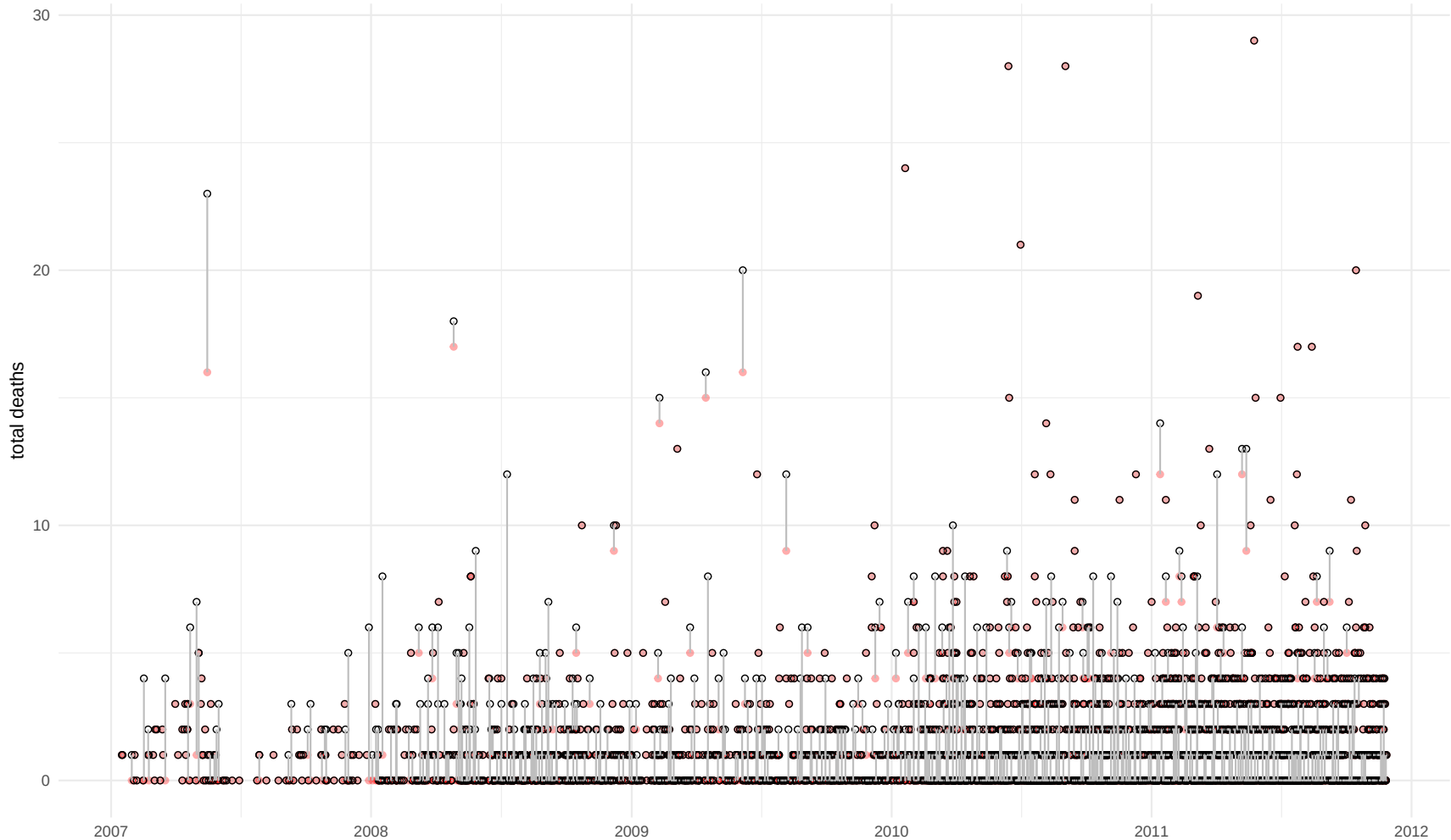
Visualizations for data exploration

a slicker way to contrast differences...



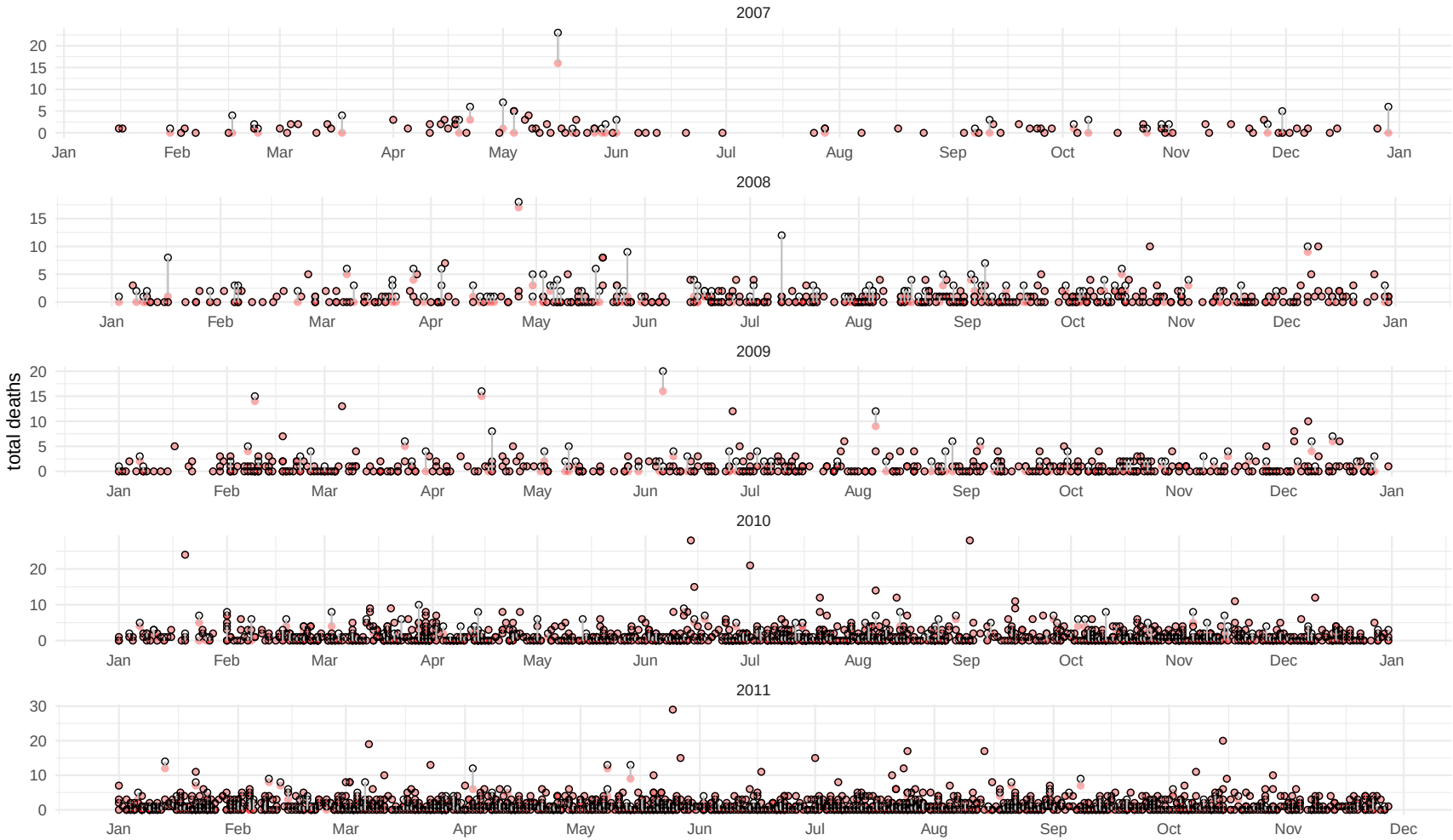
Visualizations for data exploration

let's clean up the background a bit...



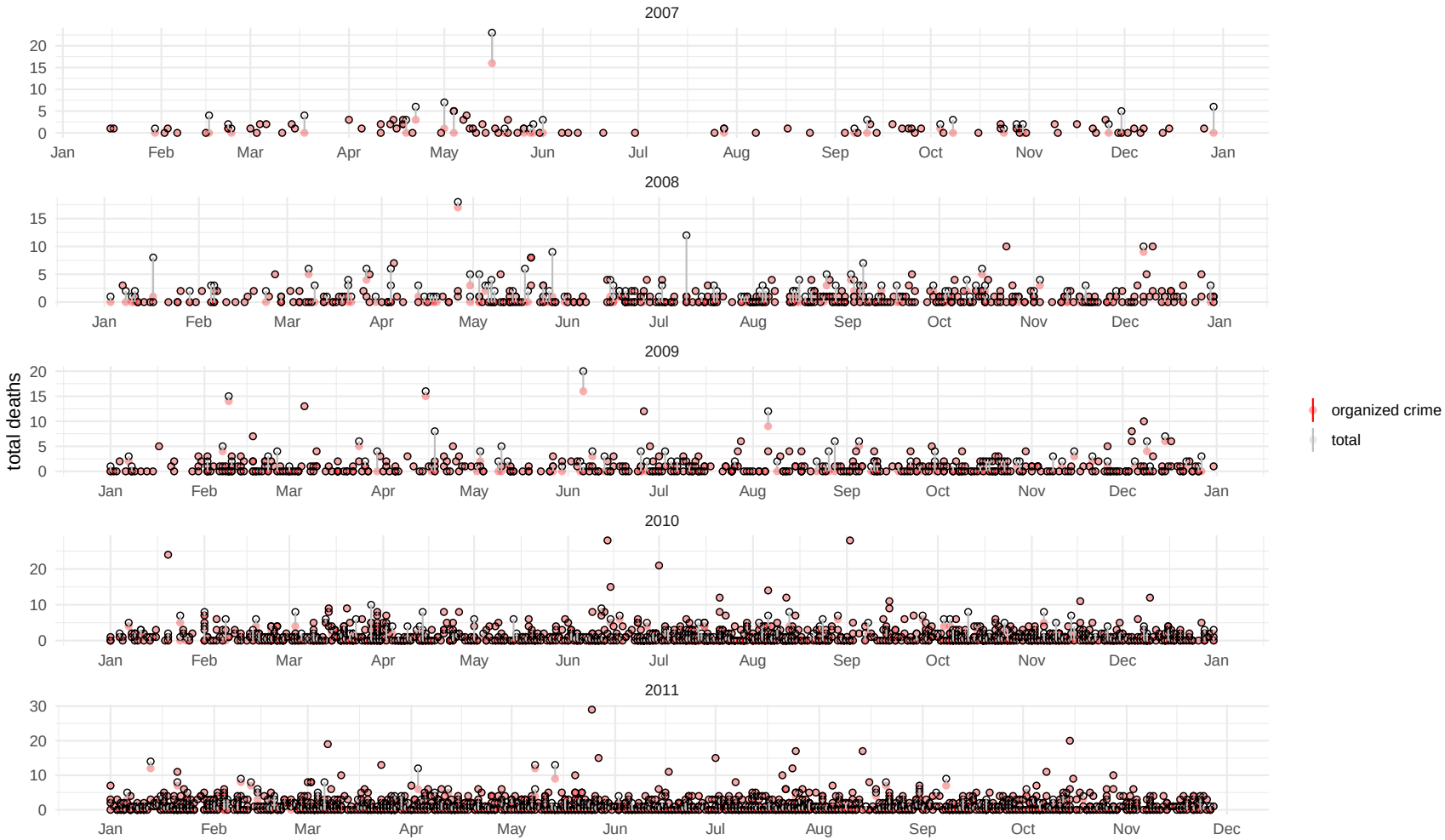
Visualizations for data exploration

maybe some patterns emerge if we break it by year..



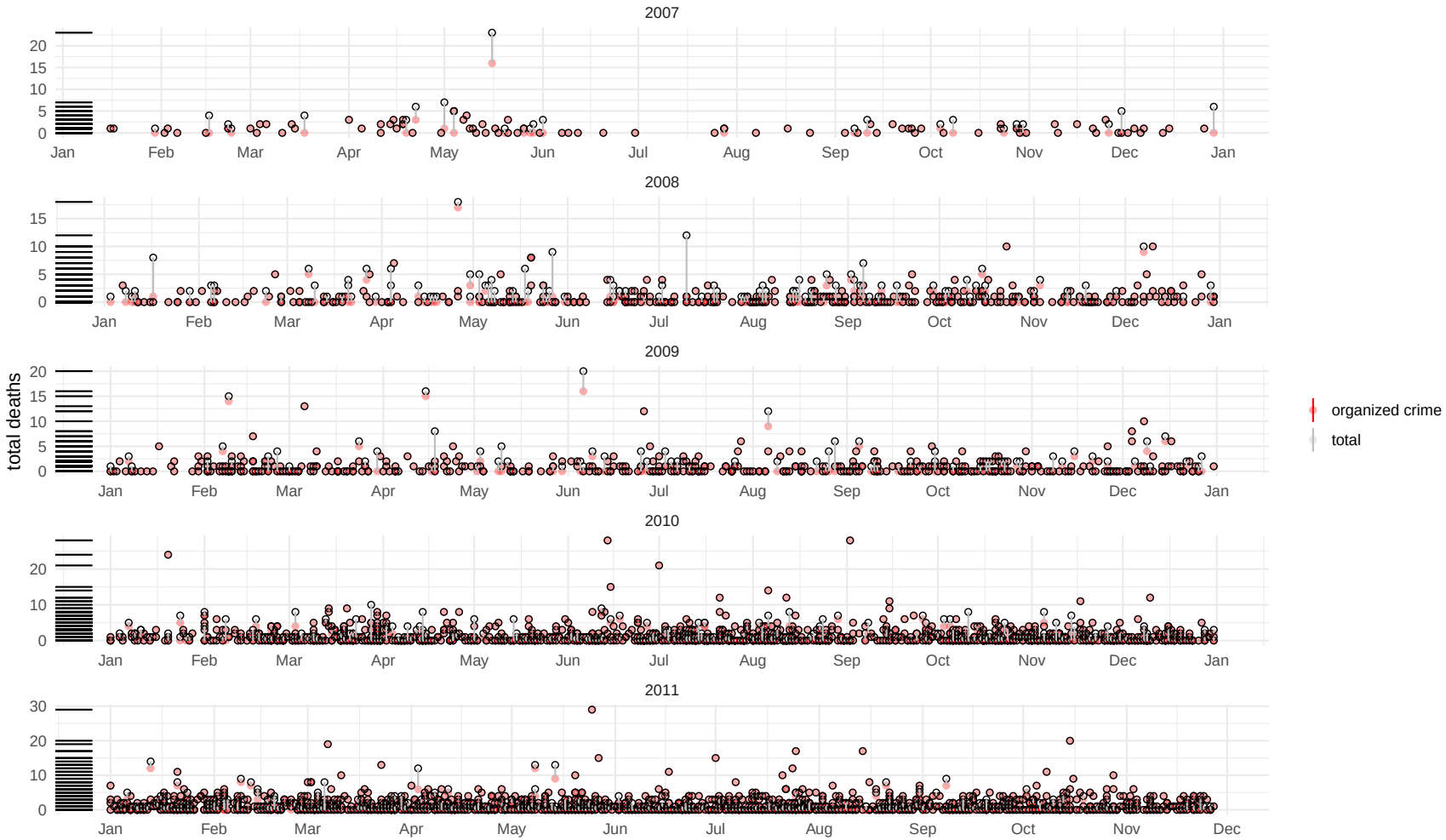
Visualizations for data exploration

what were we looking at, remind me?



Visualizations for data exploration

have we overdone it?



Visualizations for data exploration

If you bin it, put a ring on it...

Summarize results / insights

Visualizations to **summarize results / insights**

For what and how...

- **key questions** to ask before thinking of communicating results/insights
  1. **who is your audience?**
  2. **what point are you trying to make?**

- keep in mind that you may not have more than **20 secs to make a point**
  - short attention spans
  - not everyone is interested in details
  - cognitive tradeoff in audience between catching what you show and what you say

Cognitive psychology of data visualizations

Some basic rules...

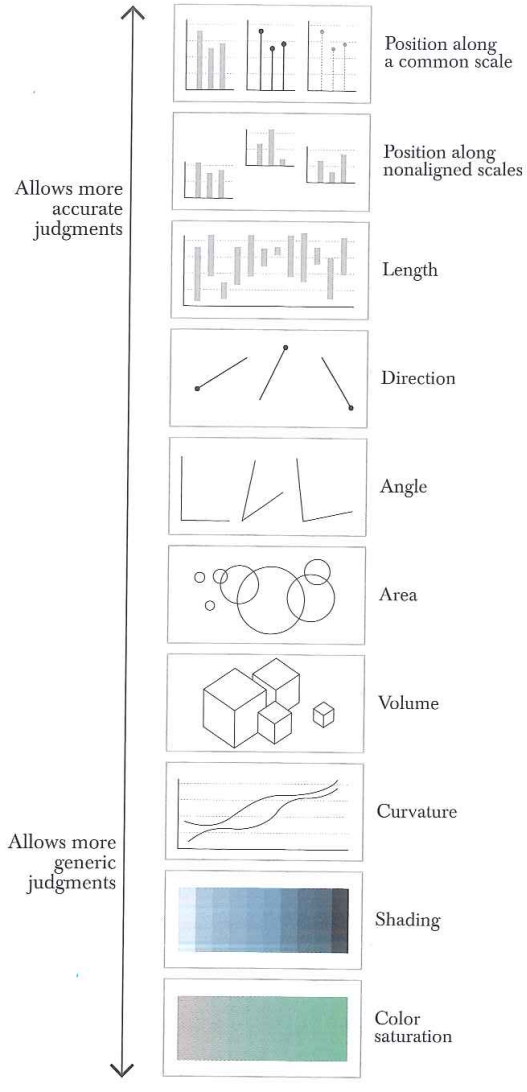
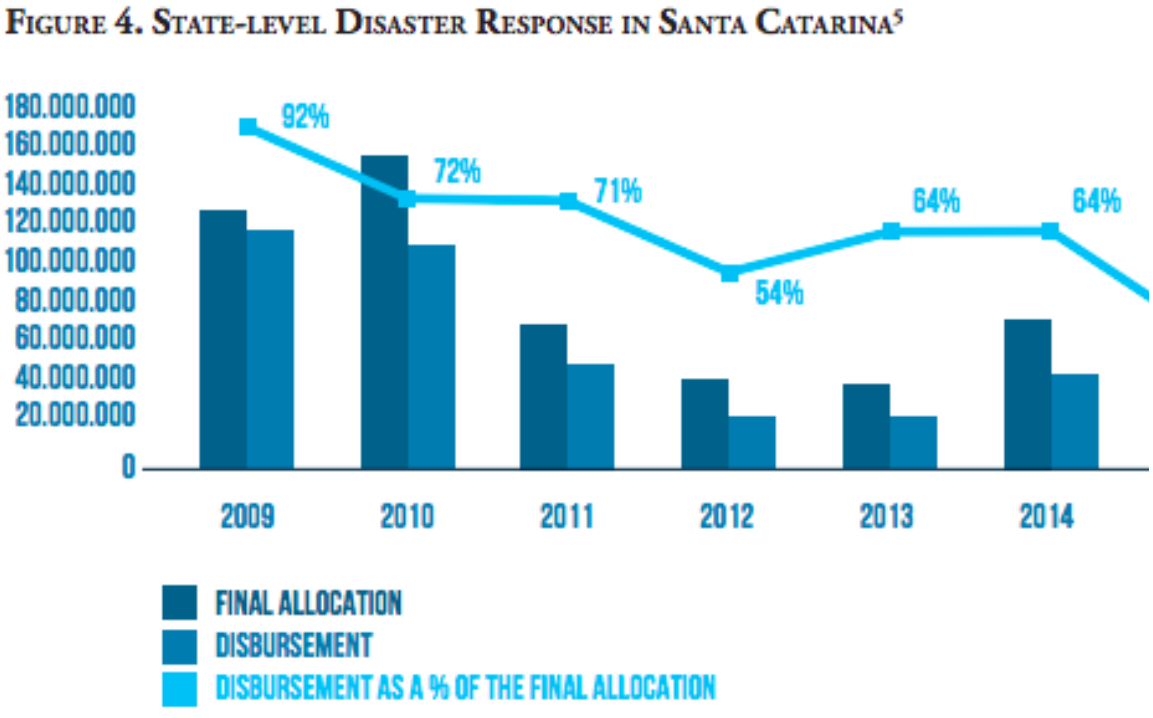
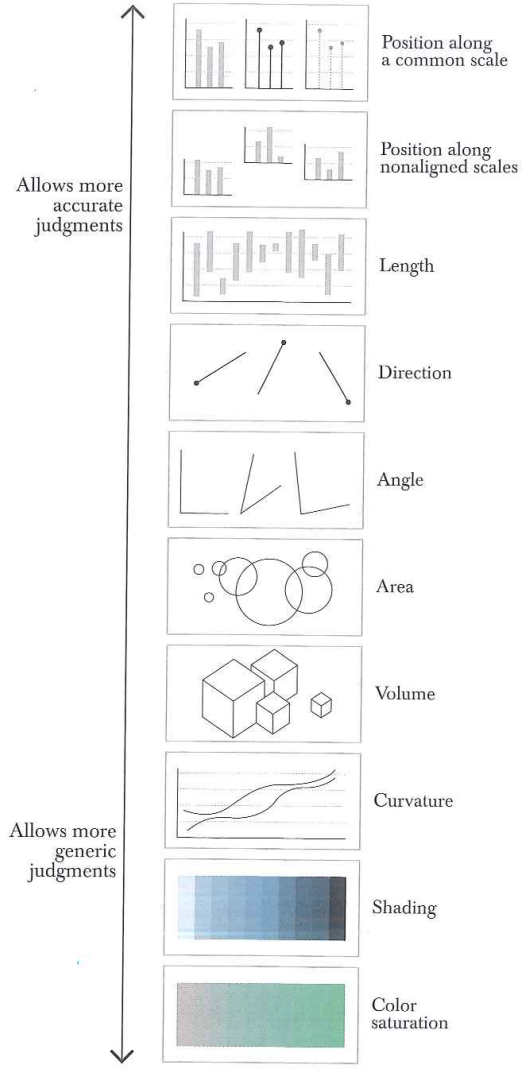


Figure 6.12 Cleveland and McGill's elementary perceptual tasks. The higher an encoding method on the scale, the more accurate the comparisons it facilitates.

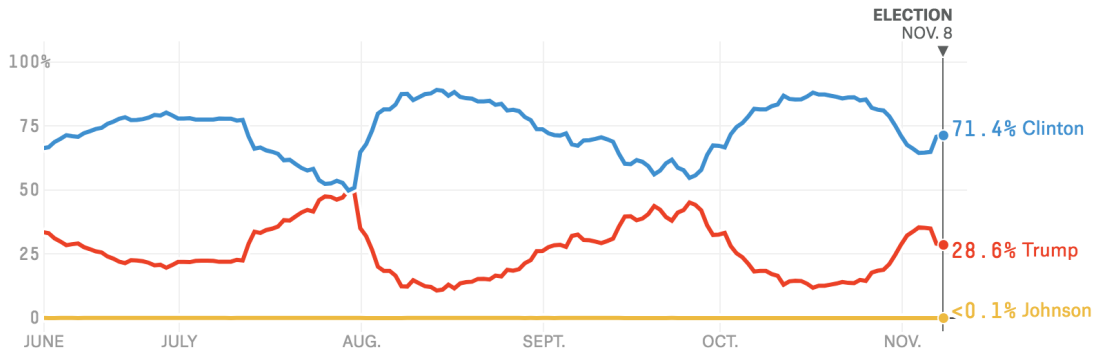
Figure: Cairo (2013)

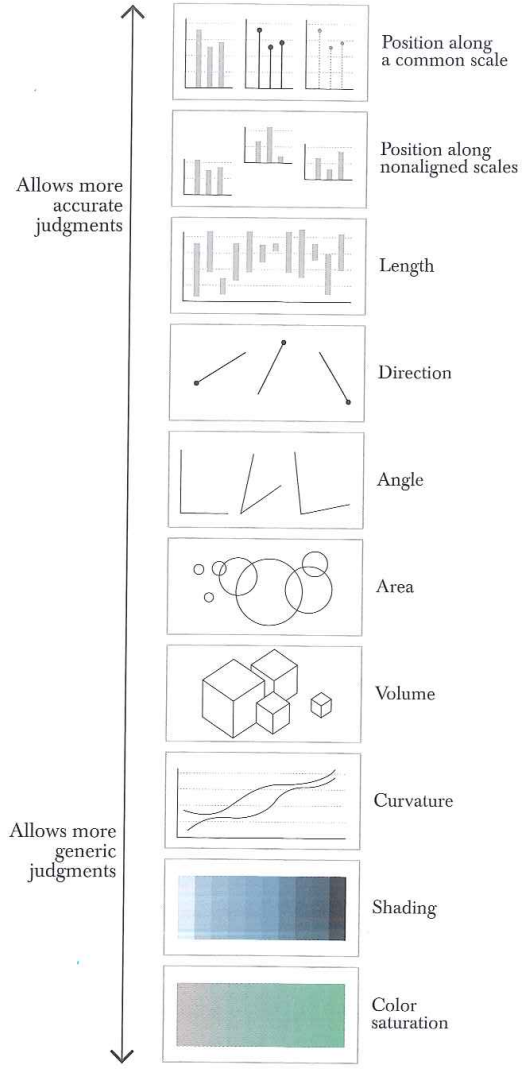
What's the cognitive effort to get this chart's insight?



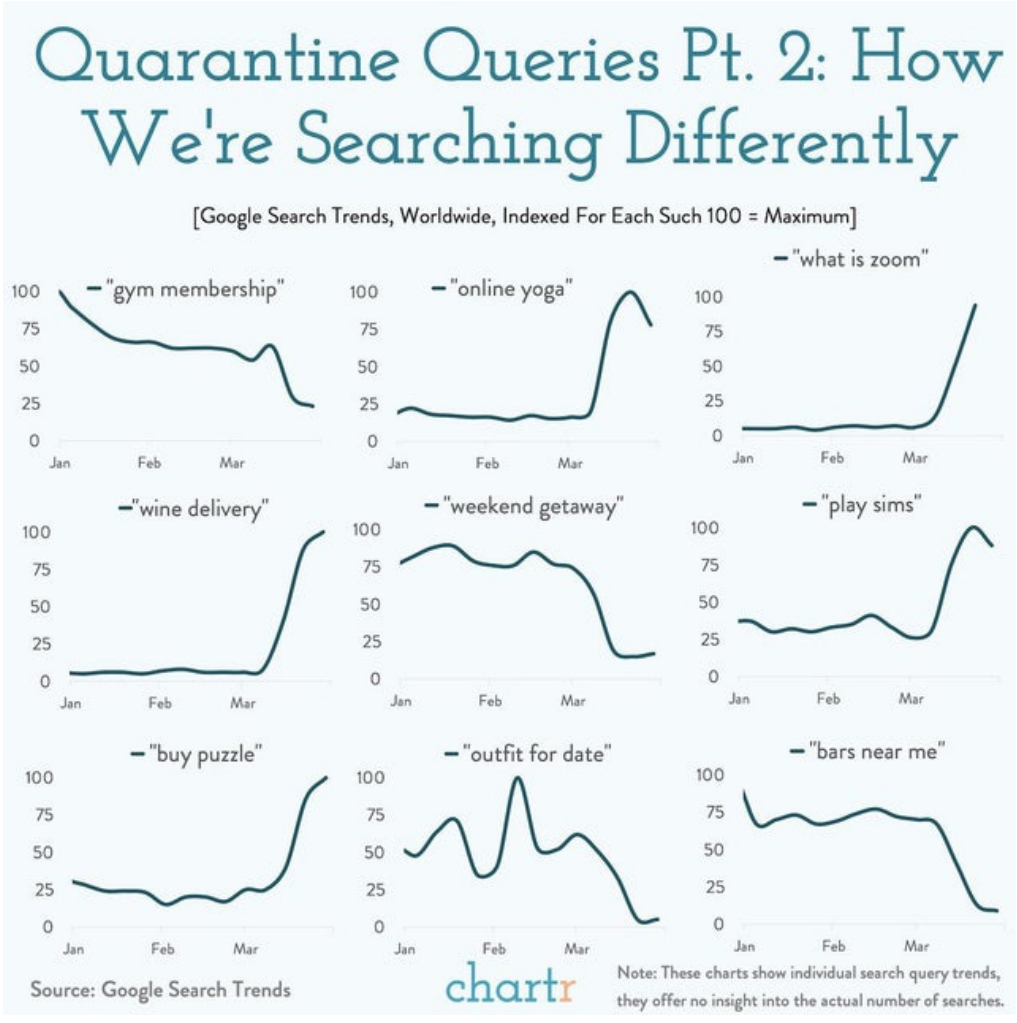


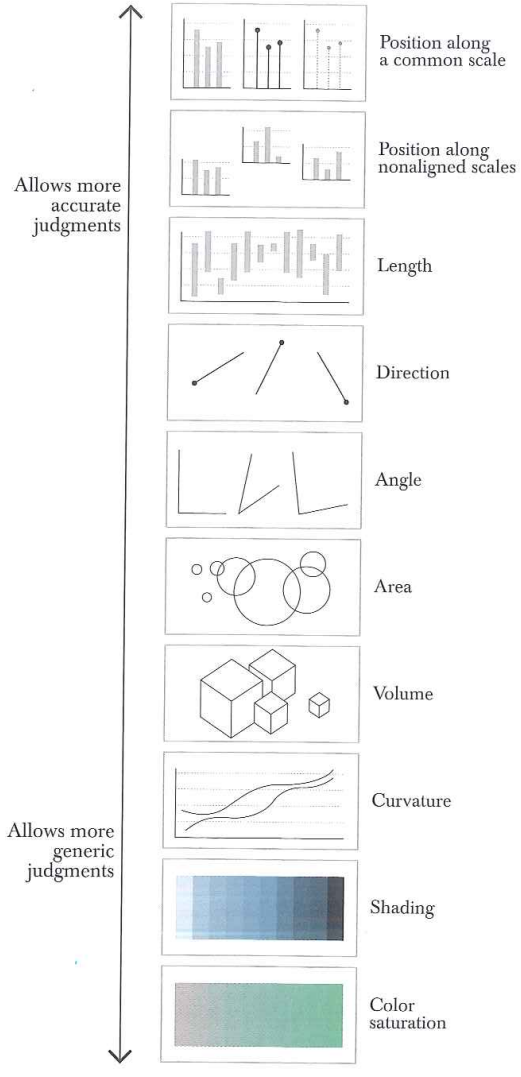
What's the cognitive effort to get this chart's insight?



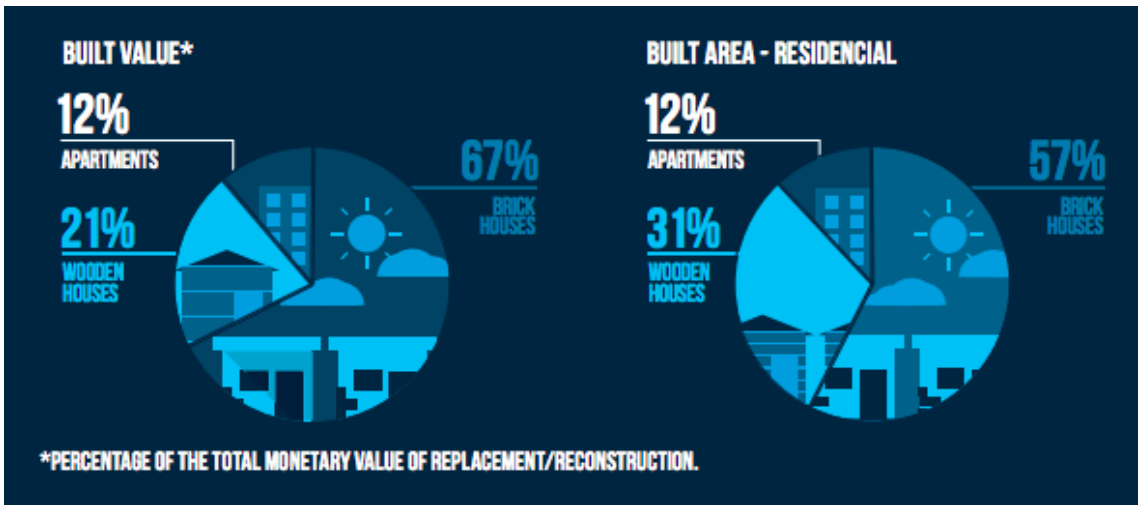


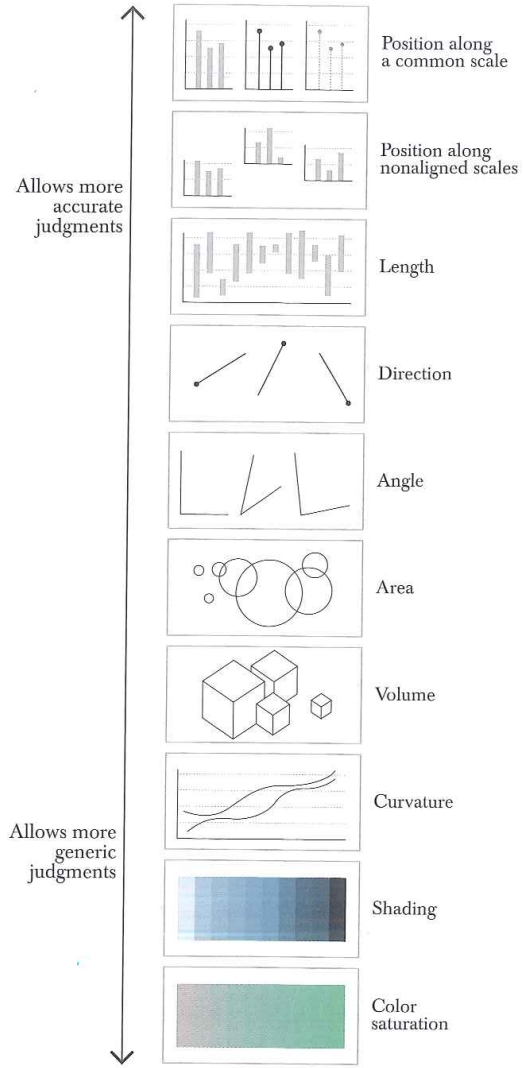
What's the cognitive effort to get this chart's insight?



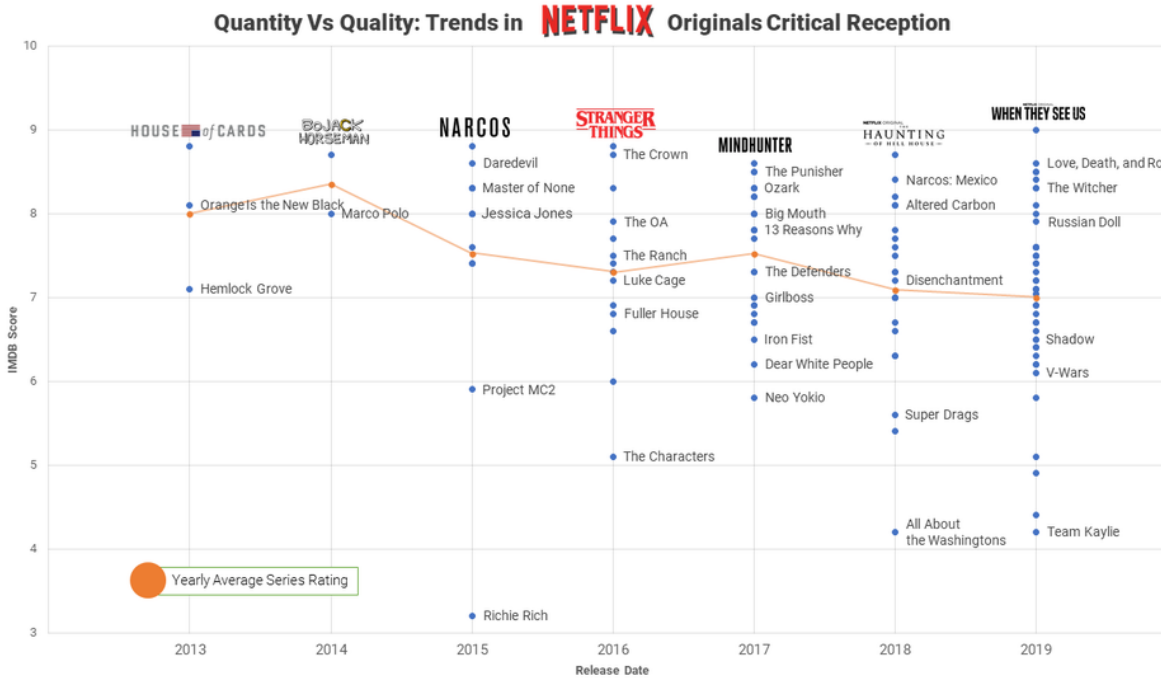


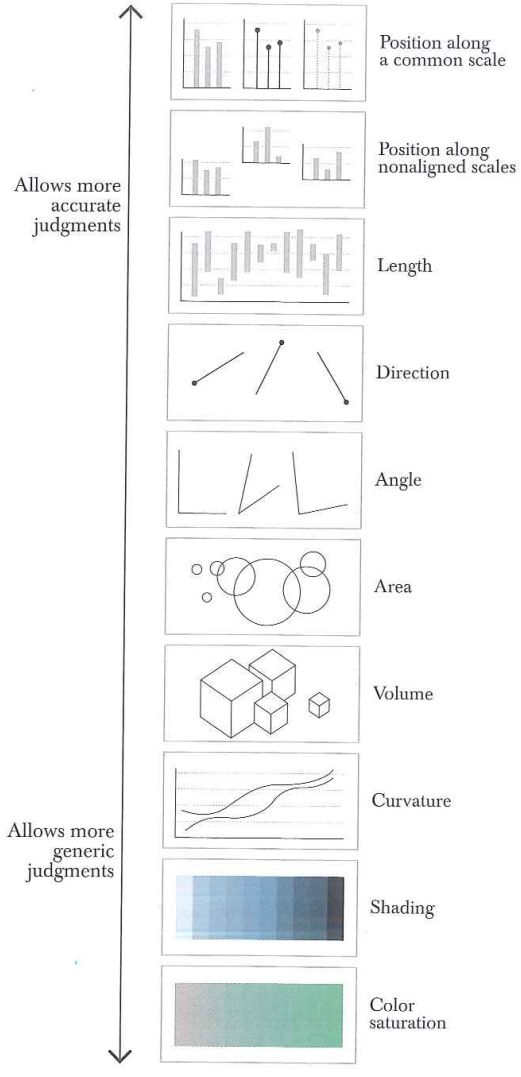
What's the cognitive effort to get this chart's insight?



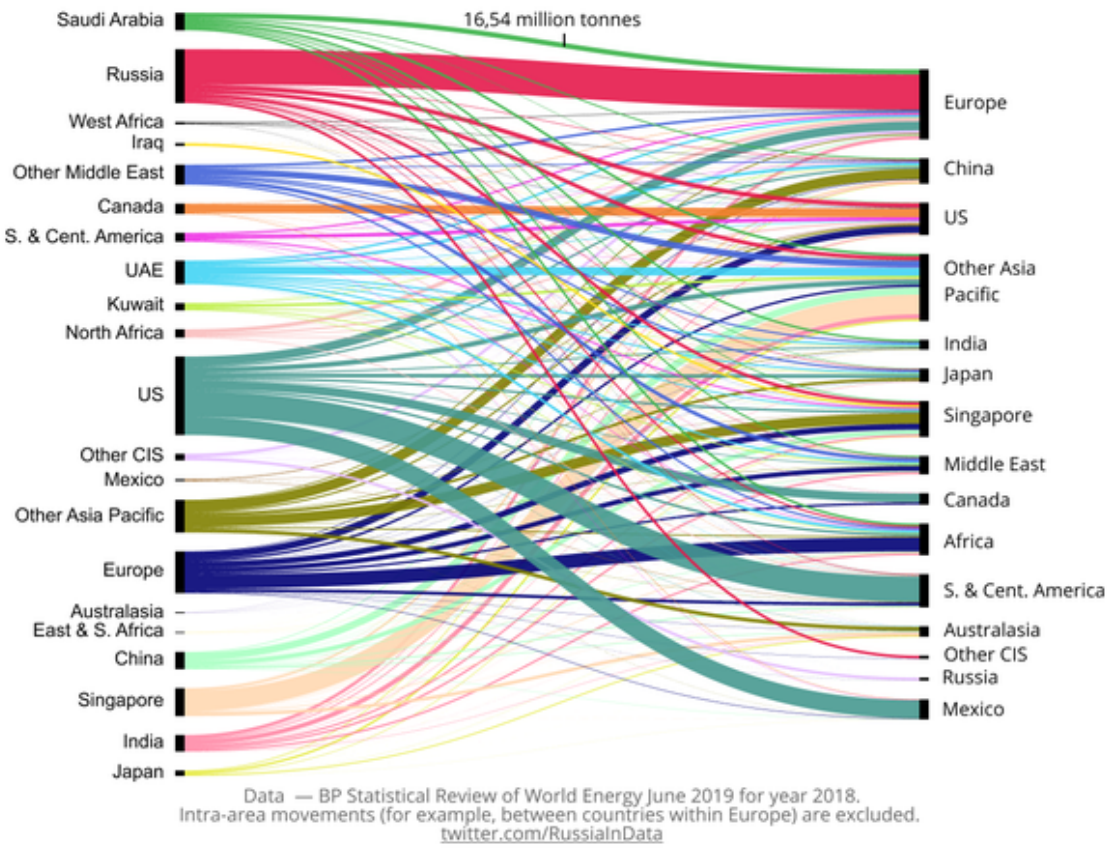


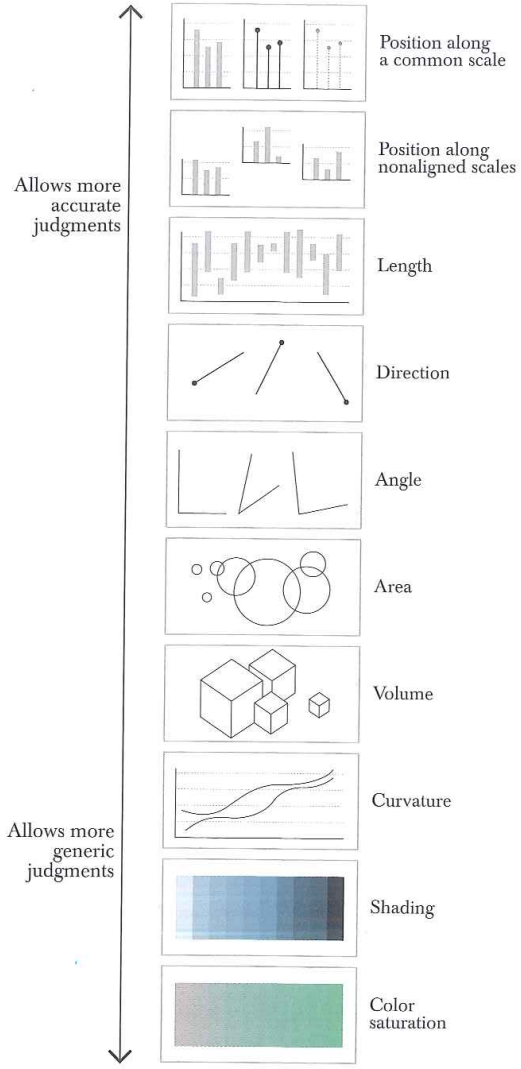
What's the cognitive effort to get this chart's insight?



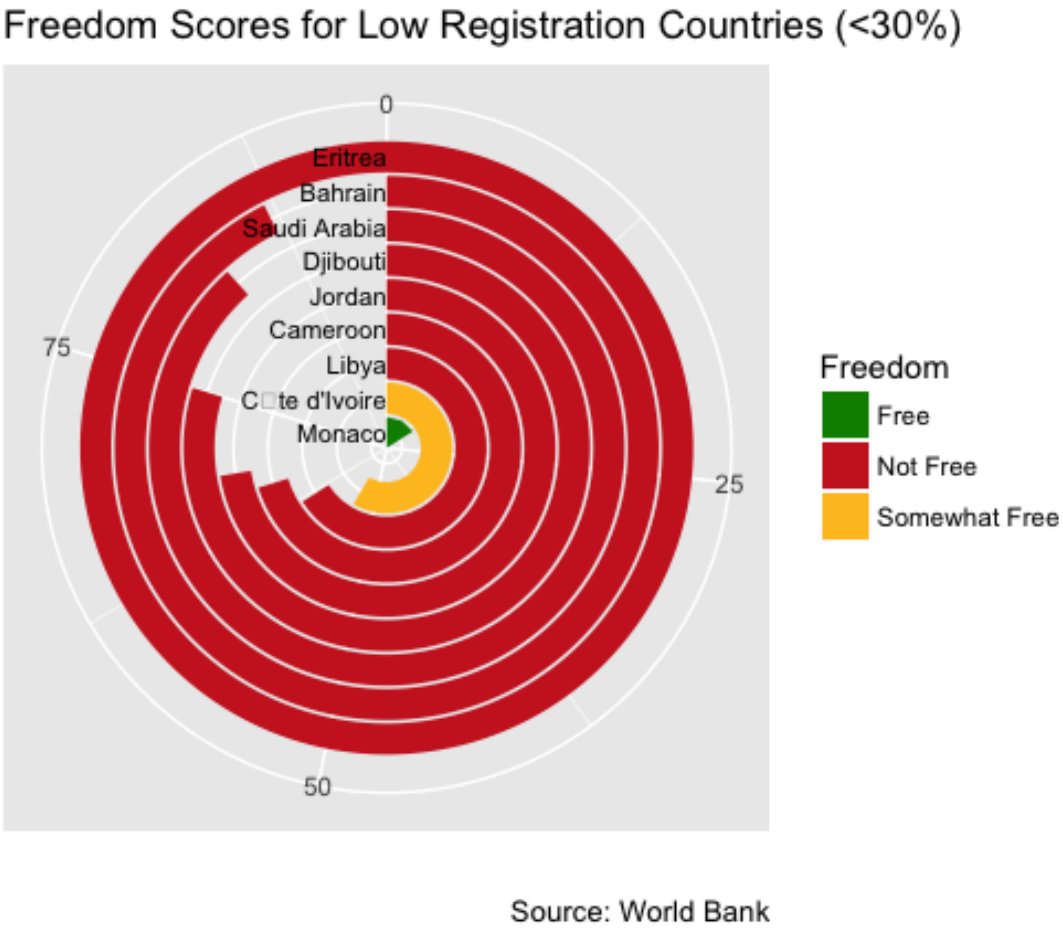


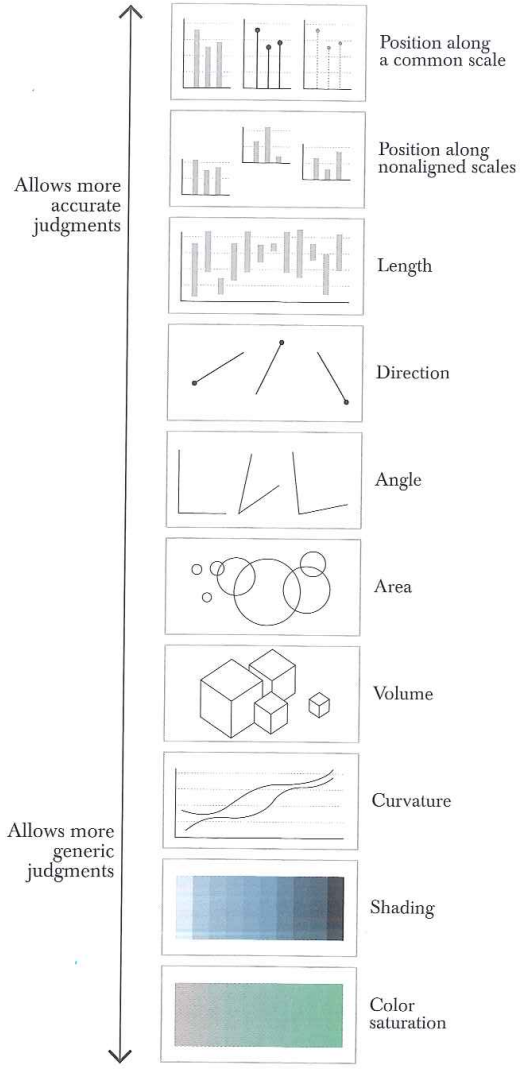
What's the cognitive effort to get this chart's insight?





What's the cognitive effort to get this chart's insight?





Visualizations for summarizing results/insights

Some basic rules...

- keep your visualization **simple**!
  - we all know you can make complicated graphs, but should you?
- use colors to **highlight the important data**
  - tone down the rest of the data (literally!)
- make **one point per graph**
- add just enough information to make it **self-explanatory**
  - careful not to de-clutter to the point of unintelligibility!
- choose the type of graph best tailored to your objectives

Data Visualizations

Marco Morales
marco.morales@columbia.edu

Nana Yaw Essuman
nanayawce@gmail.com

GR5069
Topics in Applied Data Science
for Social Scientists

Spring 2021
Columbia University