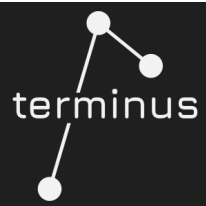


What happens in Vegas should stay in Vegas

honest storytelling with data

Derrick W. Turk
terminus data science, LLC



1. Pictures

2. Postmodernism

3. Pro-tips

4. Proscriptions

A picture is worth 1000 kB

Based on a quick Internet search, the most popular topic presented as “story-telling in data science” is *data visualization*.

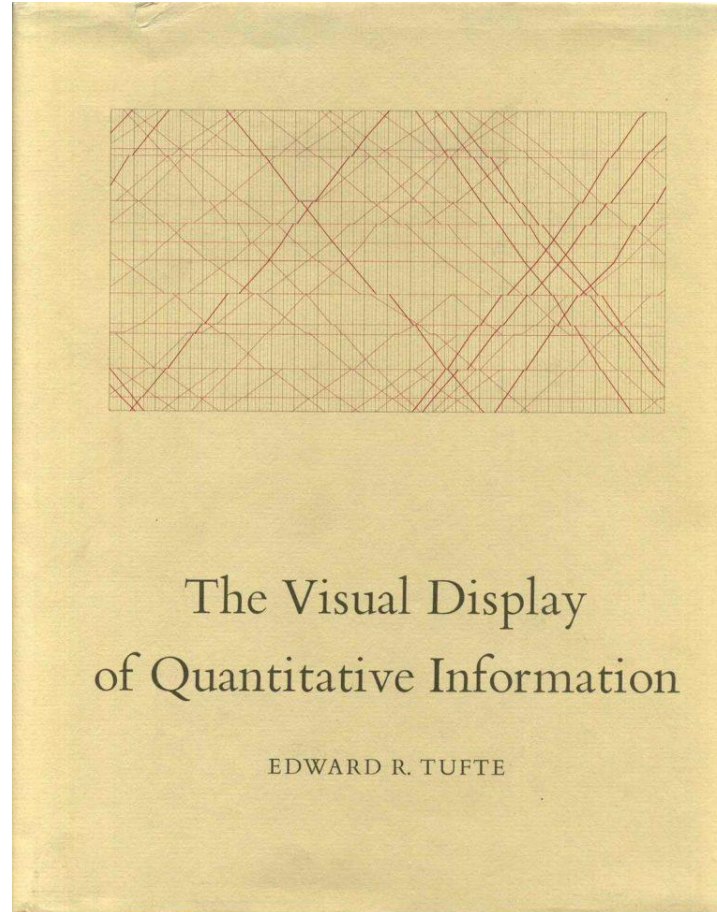
Data visualization is absolutely fantastic, and a critical part of story-telling with data

- first, because human visual processing has about 500 million years of evolution on human language processing

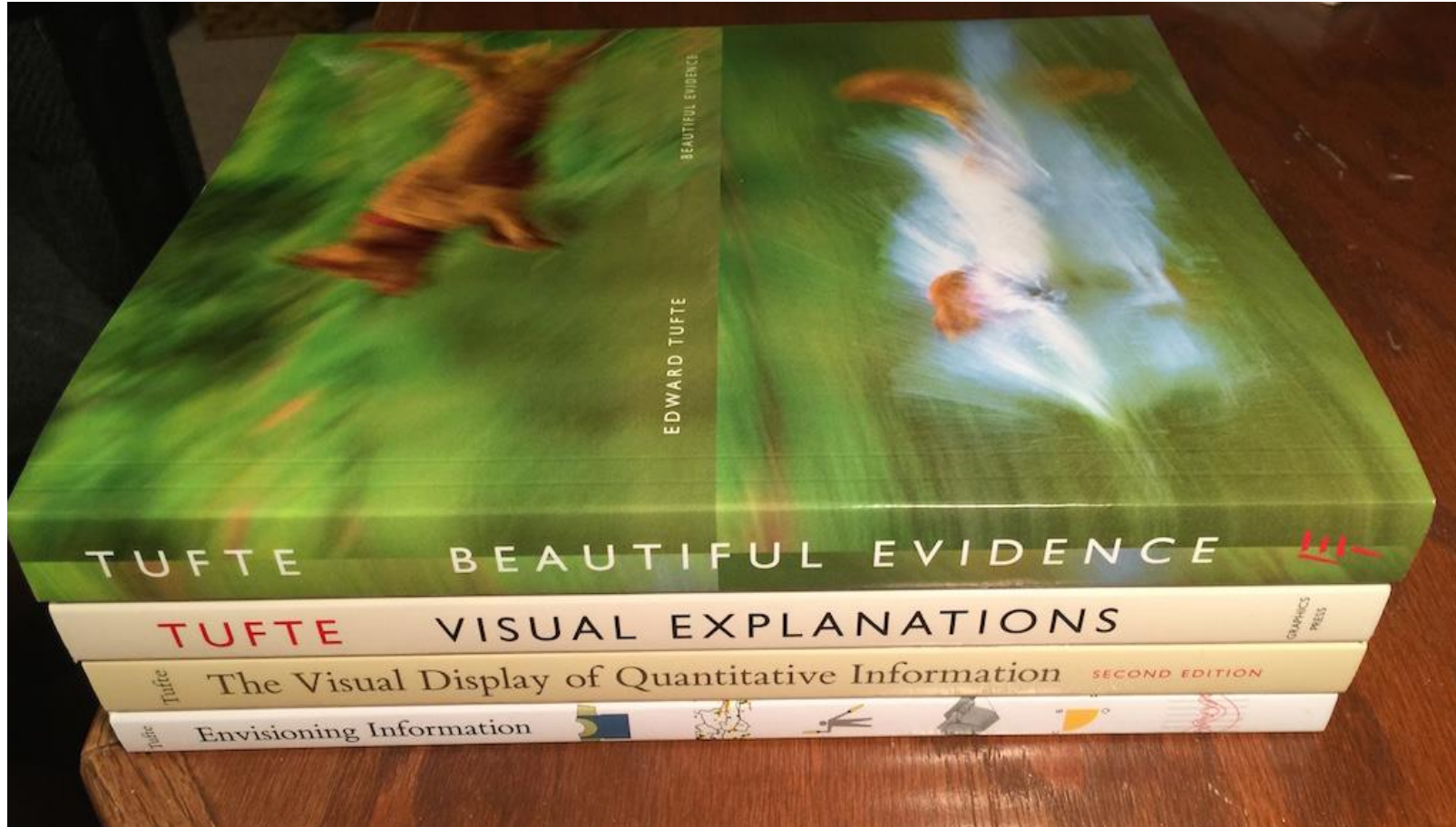
- second, because culturally we’ve come to expect *Watchmen* rather than *War and Peace Illustrated*

If you came here hoping for an in-depth study of how to effectively communicate with data visualizations...

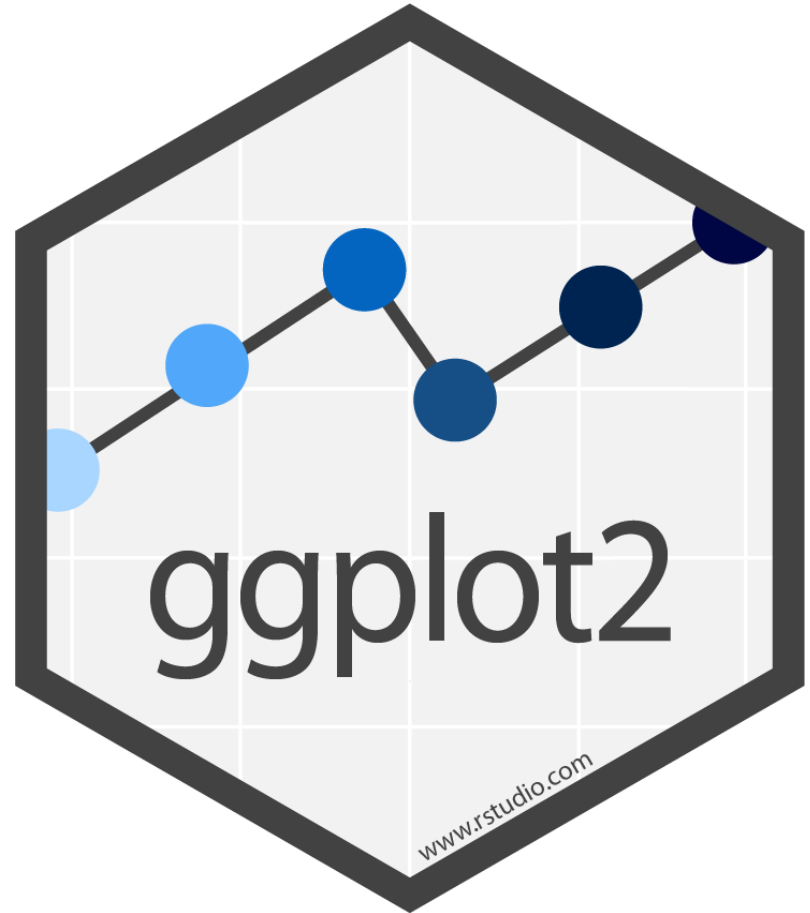
Stop watching now and read this instead



Then read these



Then learn something like these



And never be afraid to pick up these!



1. Pictures

2. Postmodernism

3. Pro-tips

4. Proscriptions

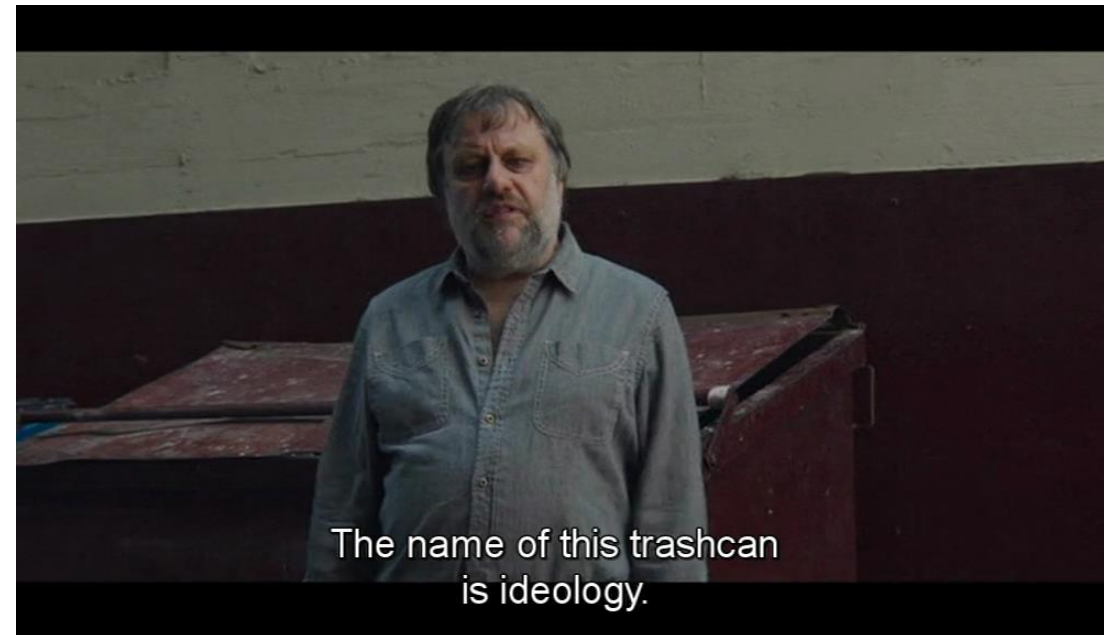
The “narrative fallacy” fallacy

In a polarized time, we can all agree on two things:

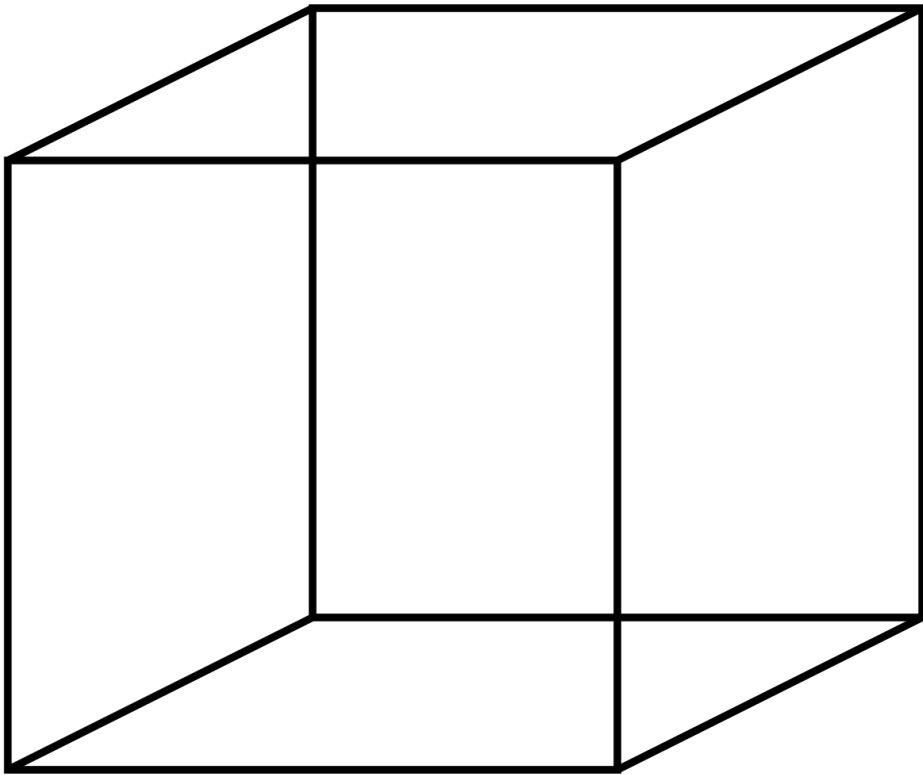
- *those other guys* are driven by an *agenda* to promote a pre-established *narrative*
- *our noble heroes* are merely reporting on *objective reality* by observing *facts*

There is no “**view from nowhere**”!

“I already am eating from the trash can all the time...”
–Slavoj Žižek, “The Pervert’s Guide to Cinema”



The view(s) from somewhere(s)



“The first principle is that **you must not fool yourself**, and you are the easiest person to fool.”
–Richard Feynman

In this house we “believe science”?

“But there is one feature I notice that is generally **missing in cargo cult science**. [...] It's a kind of scientific integrity, a principle of scientific thought that corresponds to a kind of **utter honesty**—a kind of leaning over backwards. For example, **if you're doing an experiment, you should report everything that you think might make it invalid**—not only what you think is right about it: other causes that could possibly explain your results; and things you thought of that you've eliminated by some other experiment, and how they worked—to make sure the other fellow can tell they have been eliminated.

Details that could throw doubt on your interpretation must be given, if you know them. You must do the best you can—if you know anything at all wrong, or possibly wrong—to explain it. **If you make a theory, for example, and advertise it, or put it out, then you must also put down all the facts that disagree with it, as well as those that agree with it.** There is also a more subtle problem. When you have put a lot of ideas together to make an elaborate theory, you want to make sure, when explaining what it fits, that those things it fits are not just the things that gave you the idea for the theory; but that the finished theory makes something else come out right, in addition.

In summary, the idea is to **give all of the information to help others to judge the value of your contribution**; not just the information that leads to judgement in one particular direction or another.”

—Richard Feynman, “Cargo Cult Science”

“And be not conformed to this world...”

“Dumb it down”? “More analogies”? “Too technical”?

It is possible to have *nuance* without BS: educate the audience!

“Proof by analogy is fraud” but explication by analogy is effective when done with care

If “the medium is the message” [McLuhan], what on earth sort of message are we sending with all these PowerPoint decks...?

1. Pictures

2. Postmodernism

3. Pro-tips

4. Proscriptions

Grice, Grice, baby

Obey the *Gricean maxims*:

- Quality: do not say what you believe to be false, nor that for which you lack evidence
- Quantity: make your contribution as informative as required, no more, no less
- Relevance: be relevant
- Manner: be clear



“I **used** to do drugs. I **still do**, but I **used to, too**.”

–Mitch Hedberg

Epistemic status: wild speculation

“...there are **known knowns**; there are things we know we know. We also know there are **known unknowns**; that is to say we know there are some things we do not know. But there are also **unknown unknowns**—the ones we don't know we don't know.”

–Donald Rumsfeld



Just-in-time teaching

As a data scientist, you're also a *science communicator*, which means that you must be a **teacher**

Never be afraid to sidetrack to explain a key concept: **misconceptions** are easy to address in a working session, and very hard once they've made it into IR slide decks!

Less



, more



1. Pictures

2. Postmodernism

3. Pro-tips

4. Proscriptions



Vegas sin #1

misleading framing and omitted context

Many quantities have a natural “zero”; graphs which ignore this should **raise suspicion**



Vegas sin #2

focusing on observations over distributions



The **last ten spins** of the wheel don't matter!

The initial production of the **best well** doesn't matter!

The initial production of the **average well** matters only a little more

Show us the **distribution**!

Vegas sin #3

inappropriately-conditional probabilities

It's easy to mislead by sharing $P(A|B)$ and conveniently leaving out $P(B)$!

And beware the related treason of the *weaponized null hypothesis*:
“there is no evidence that parachutes prevent injuries from skydiving”

(This also happens in the form of *net* numbers where *gross* numbers would be more useful.)



Thanks, goodbye

