



Data Visualization and Visual Analytics Storytelling

Study Program Data Science
Prof. Dr. Tillmann Schwörer

Lecture Roadmap

Data Domains

Comparing Categories | Relationships | Geospatial | Time |
Part-to-whole | Distributions | Uncertainty | ...

Storytelling

Perception +
Visualization Design

Python + Tools

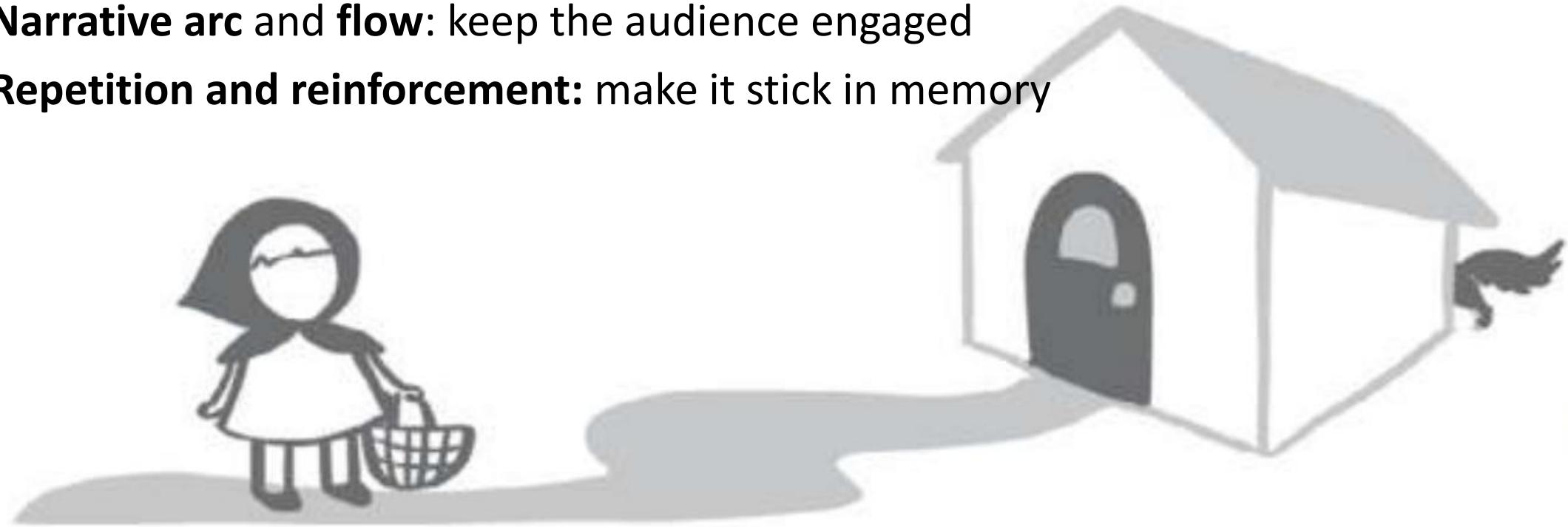
Interactive
Visualization

What makes a good story?



What makes a good story?

- ▶ **Conflict and tension:** good stories start with a problem, challenge, surprise, etc
- ▶ **Narrative arc and flow:** keep the audience engaged
- ▶ **Repetition and reinforcement:** make it stick in memory



Archetypical stories in movies



What can we learn from such movies?



Conflict and Tension

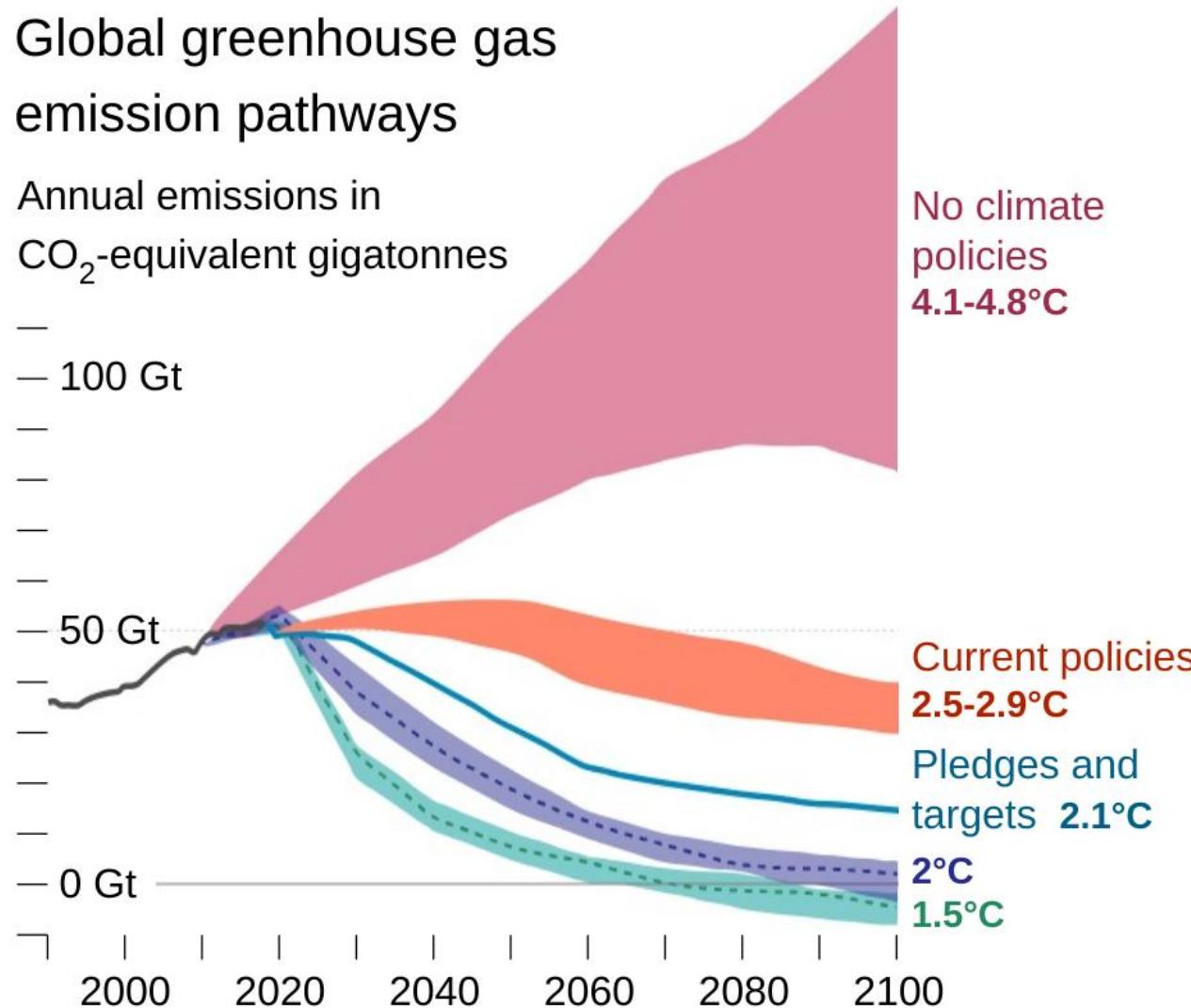
When conceptionalizing a (visual) story, reflect:

- ▶ Is there some **conflict or tension**?
- ▶ What is **at stake**? For **whom**?
- ▶ What can be **done about it**? What can the **audience** do?

Data Science is all about using data and algorithms to inform/improve decision making

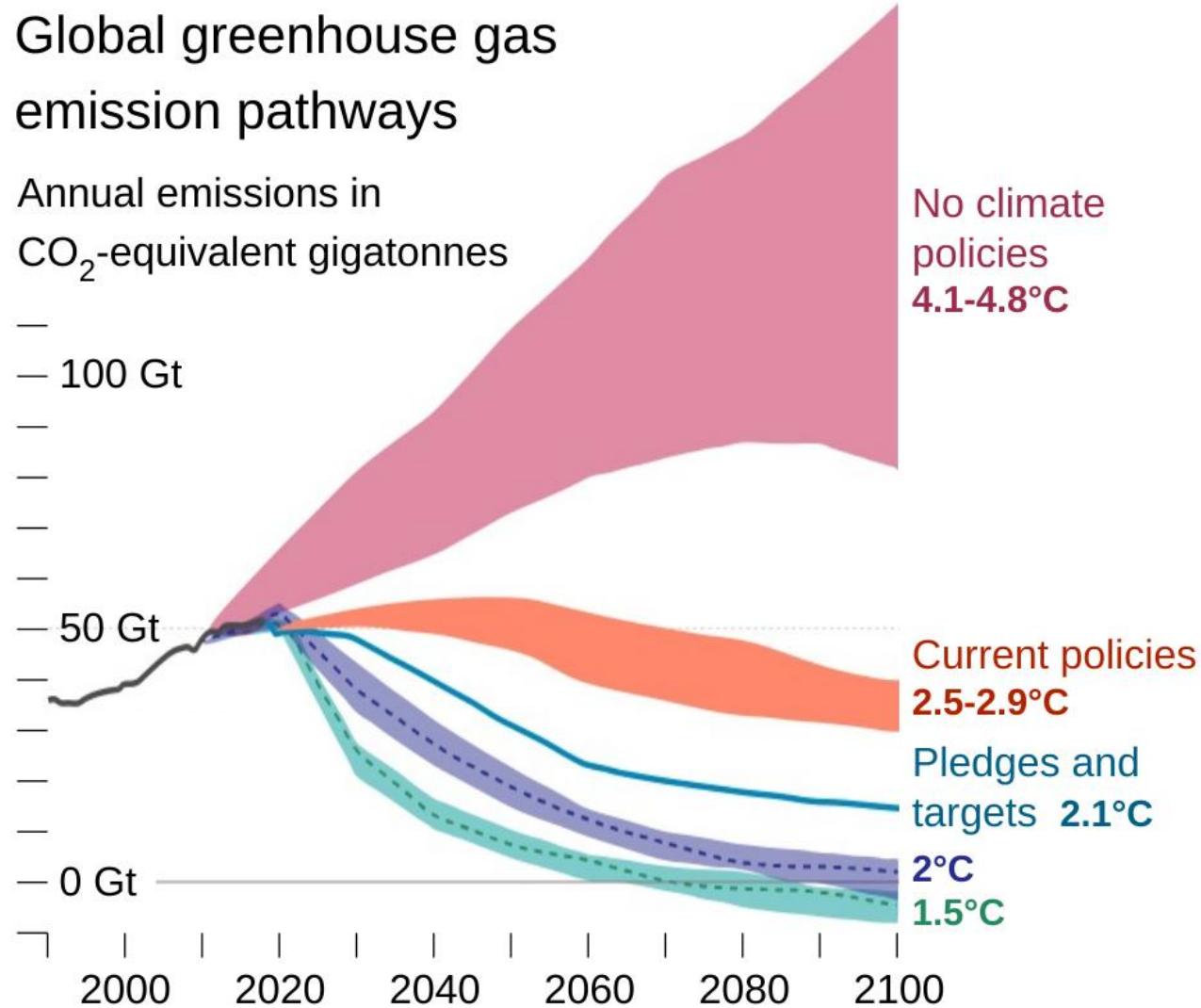
- ▶ If nothing follows from a visualization – if no decisions or actions are addressed – then maybe we should drop this visualization

1. What's the story?



- ▶ Likely disaster for many/all
- ▶ We see it coming
- ▶ How can we stop it?
 - ◆ Political decisions
 - ◆ Individual behaviour
 - ◆ Technological innovation

1. Imminent threat

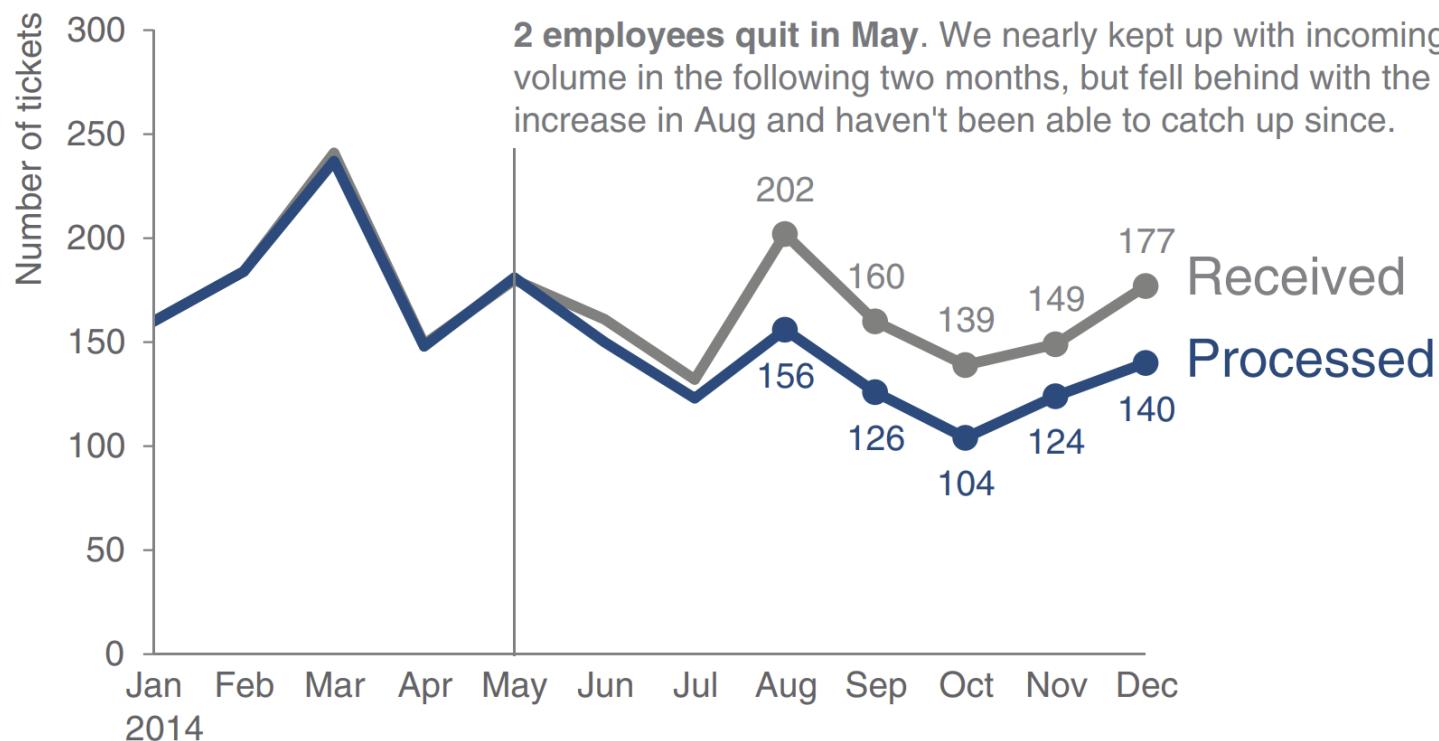


2. What's the story?

Please approve the hire of 2 FTEs

to backfill those who quit in the past year

Ticket volume over time



Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

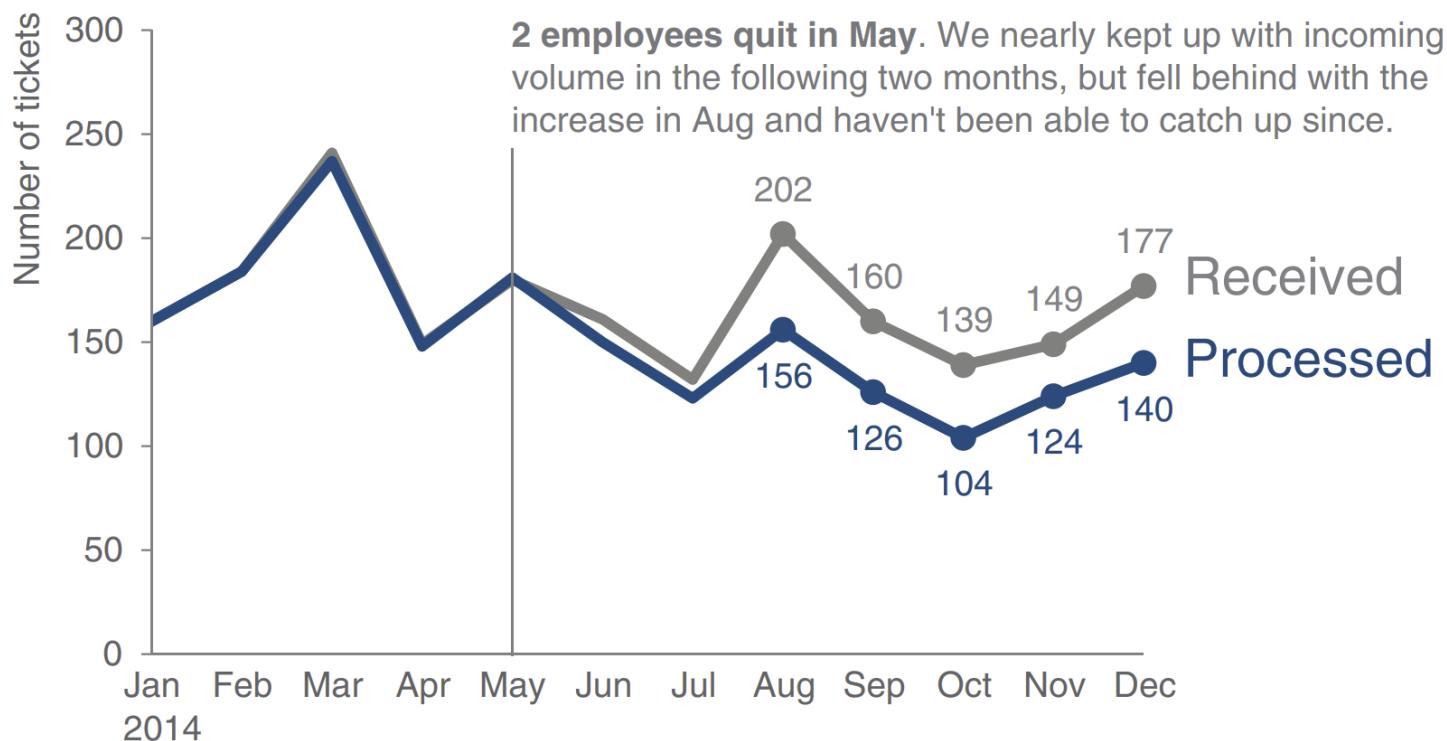
- ▶ Trigger event or decision
- ▶ Things have turned bad
- ▶ How can we go back to normal?

2. Events or decisions shift the course of history

Please approve the hire of 2 FTEs

to backfill those who quit in the past year

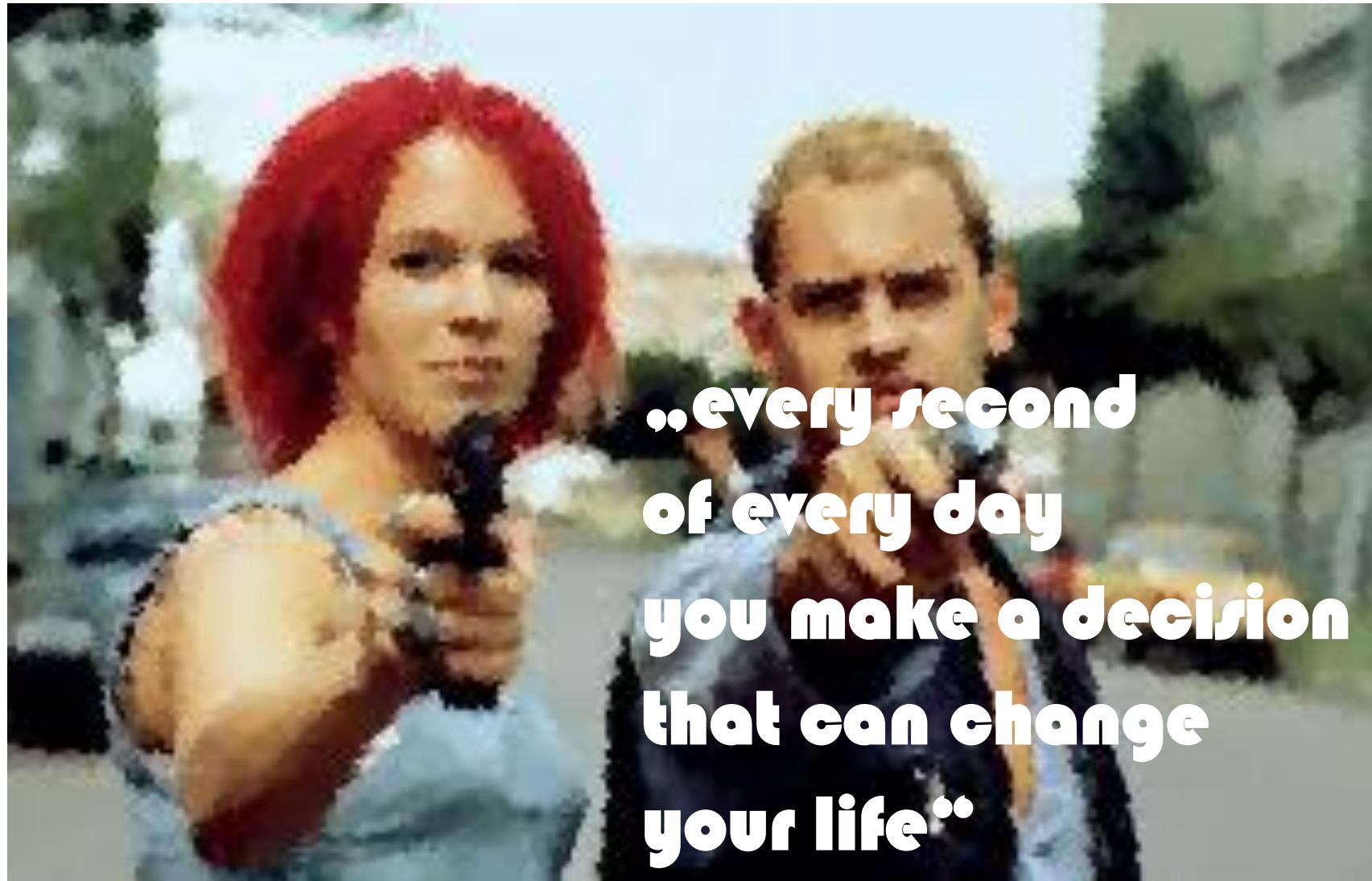
Ticket volume over time



Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

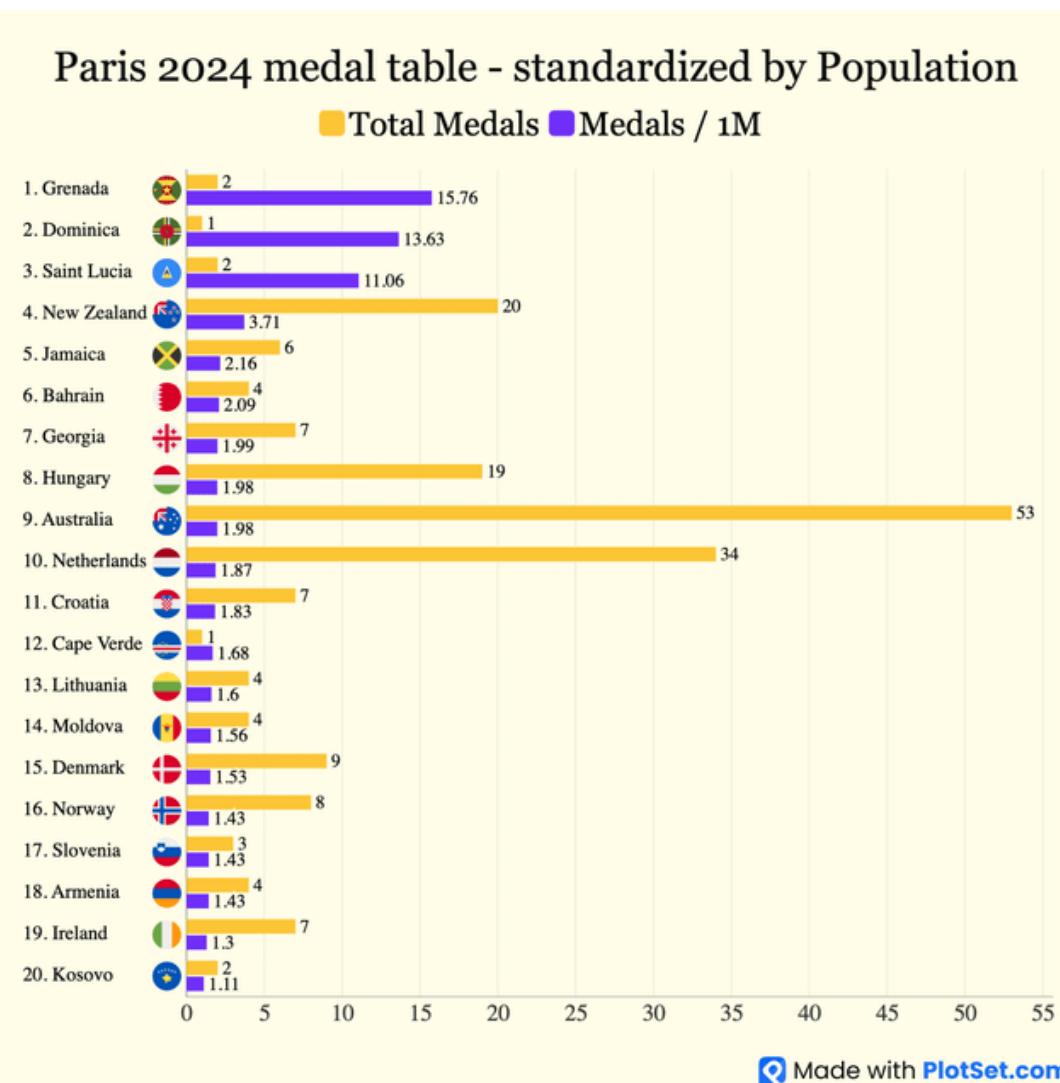


2. Events or decisions shift the course of history



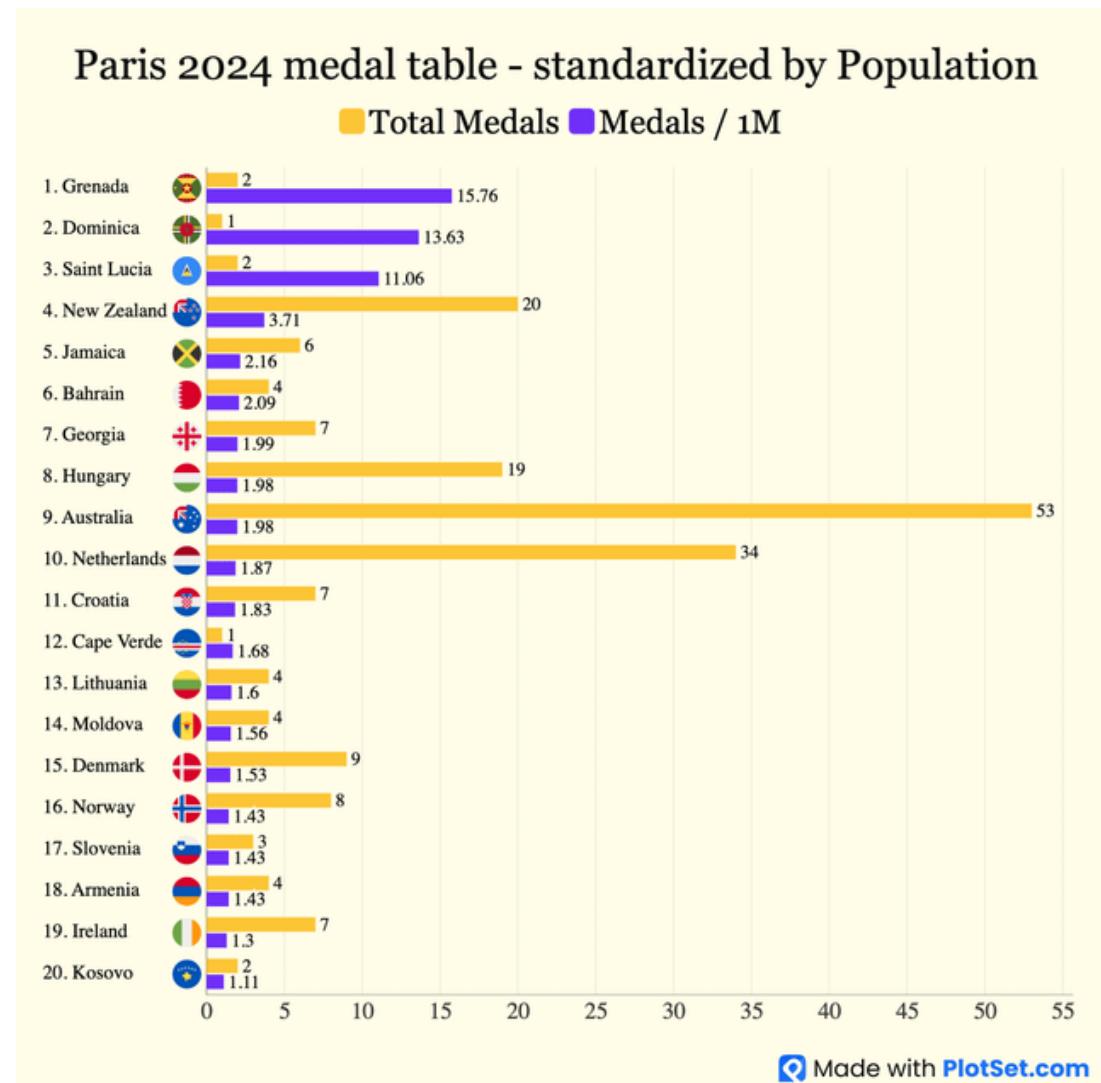
**„every second
of every day
you make a decision
that can change
your life“**

3. What is the story here?

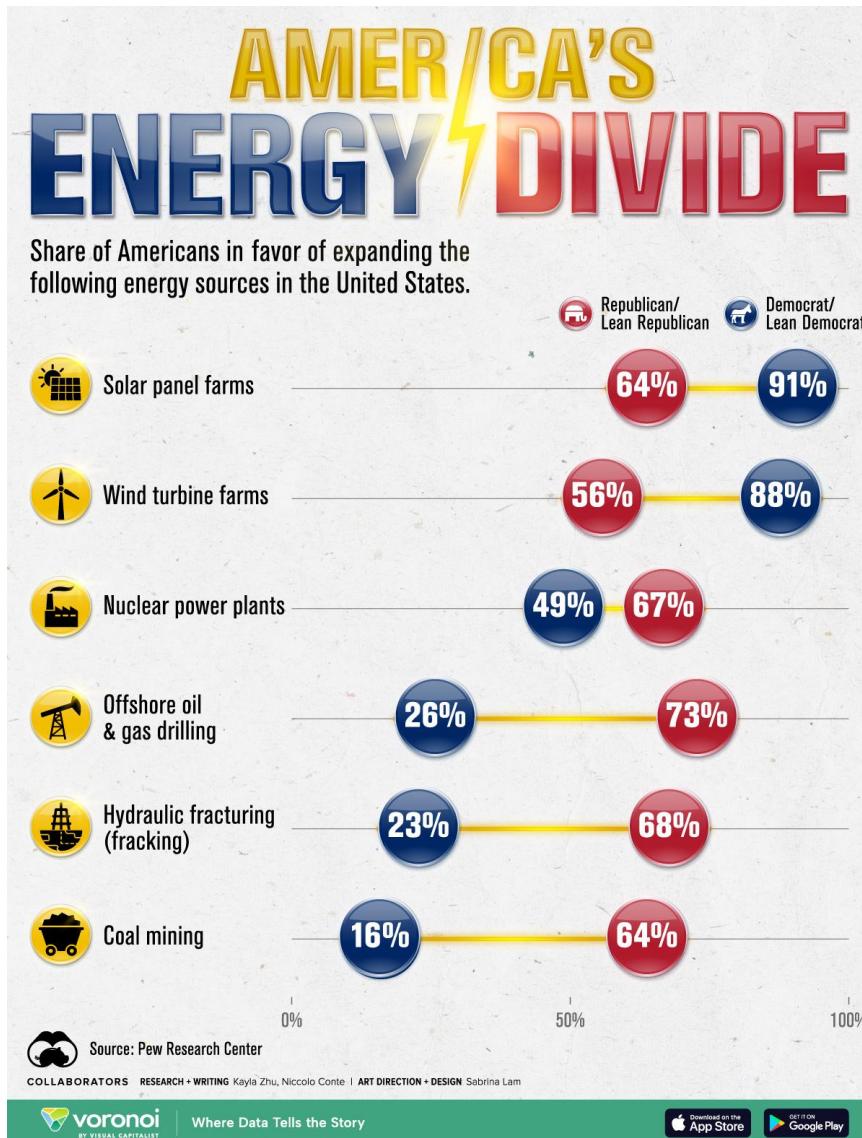


- ▶ Hopeless endeavour?
- ▶ Will the small countries have any chance?
- ▶ Can the small countries surprise?

3. Small against large

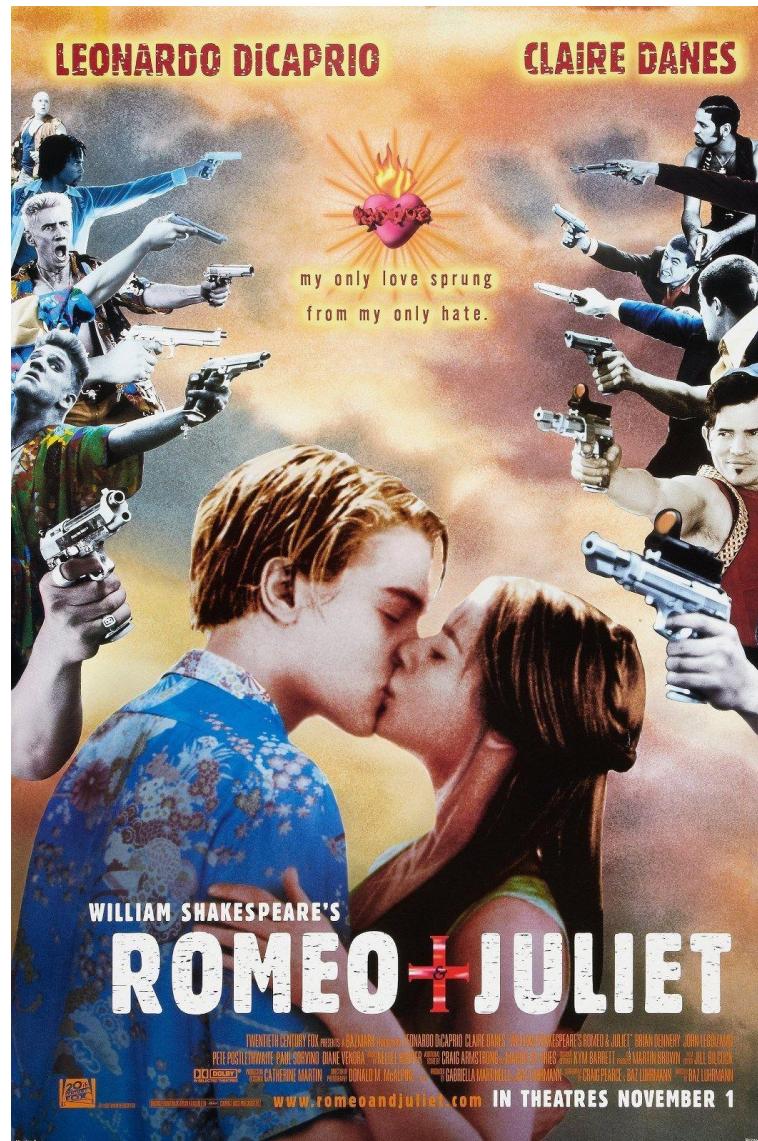
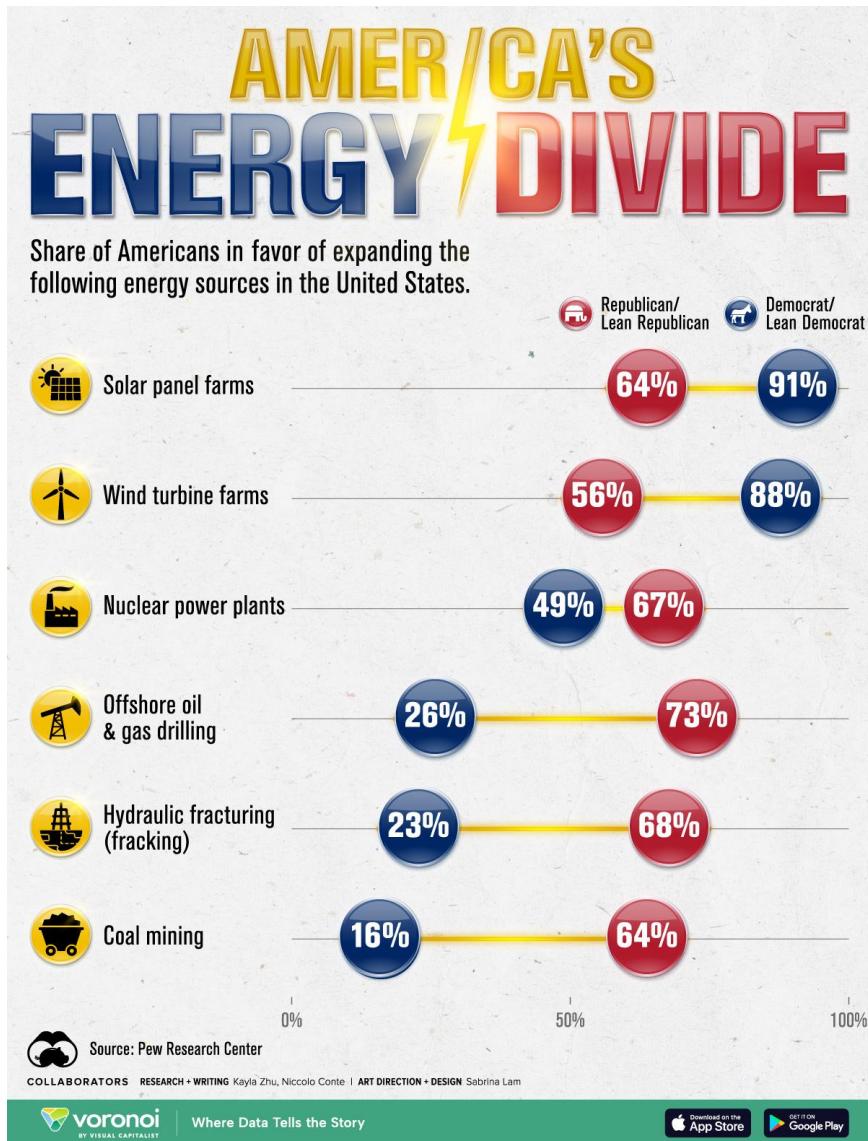


4. What is the story here?

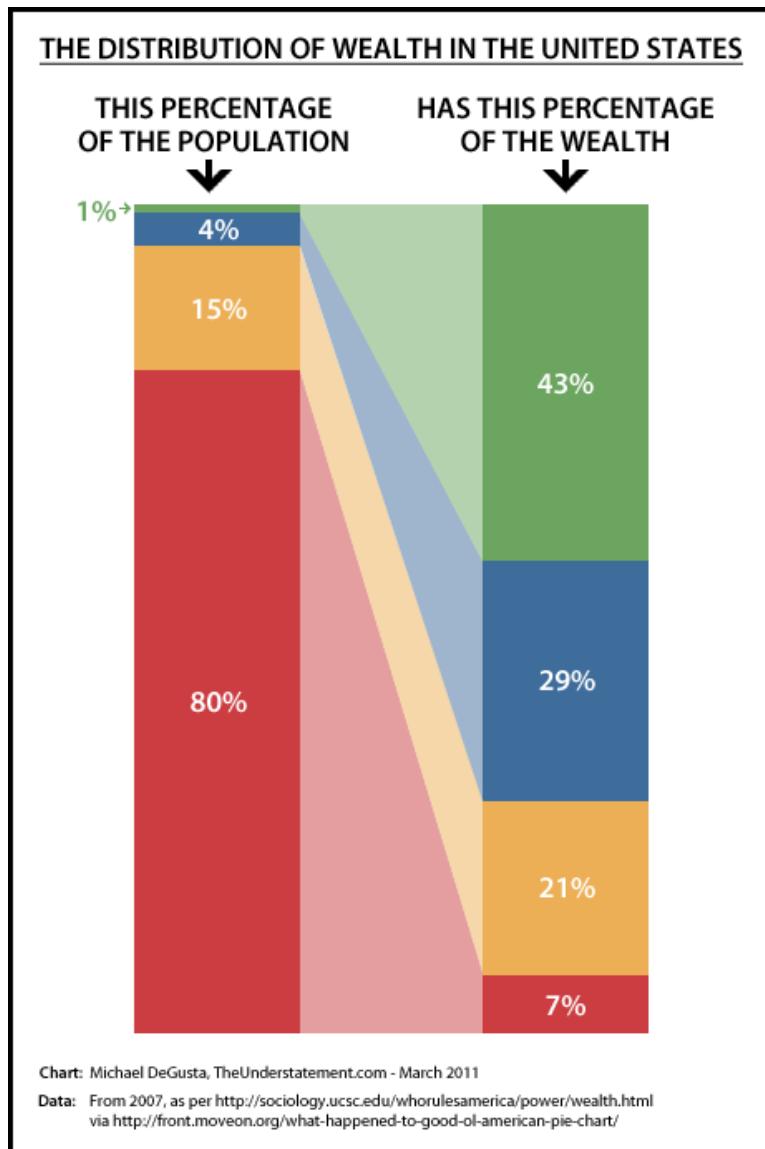


- ▶ Two antagonistic groups
- ▶ Can they come together?
- ▶ Or will there be a tragic end?

4. Impossible love

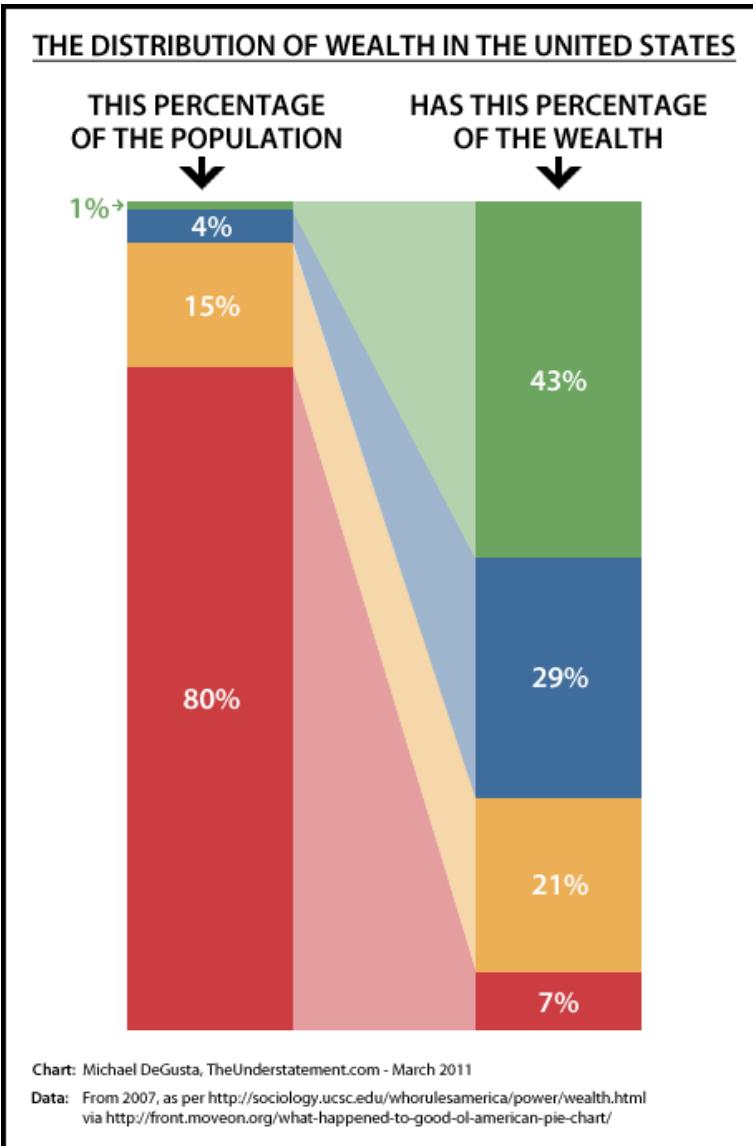


5. What is the story here?



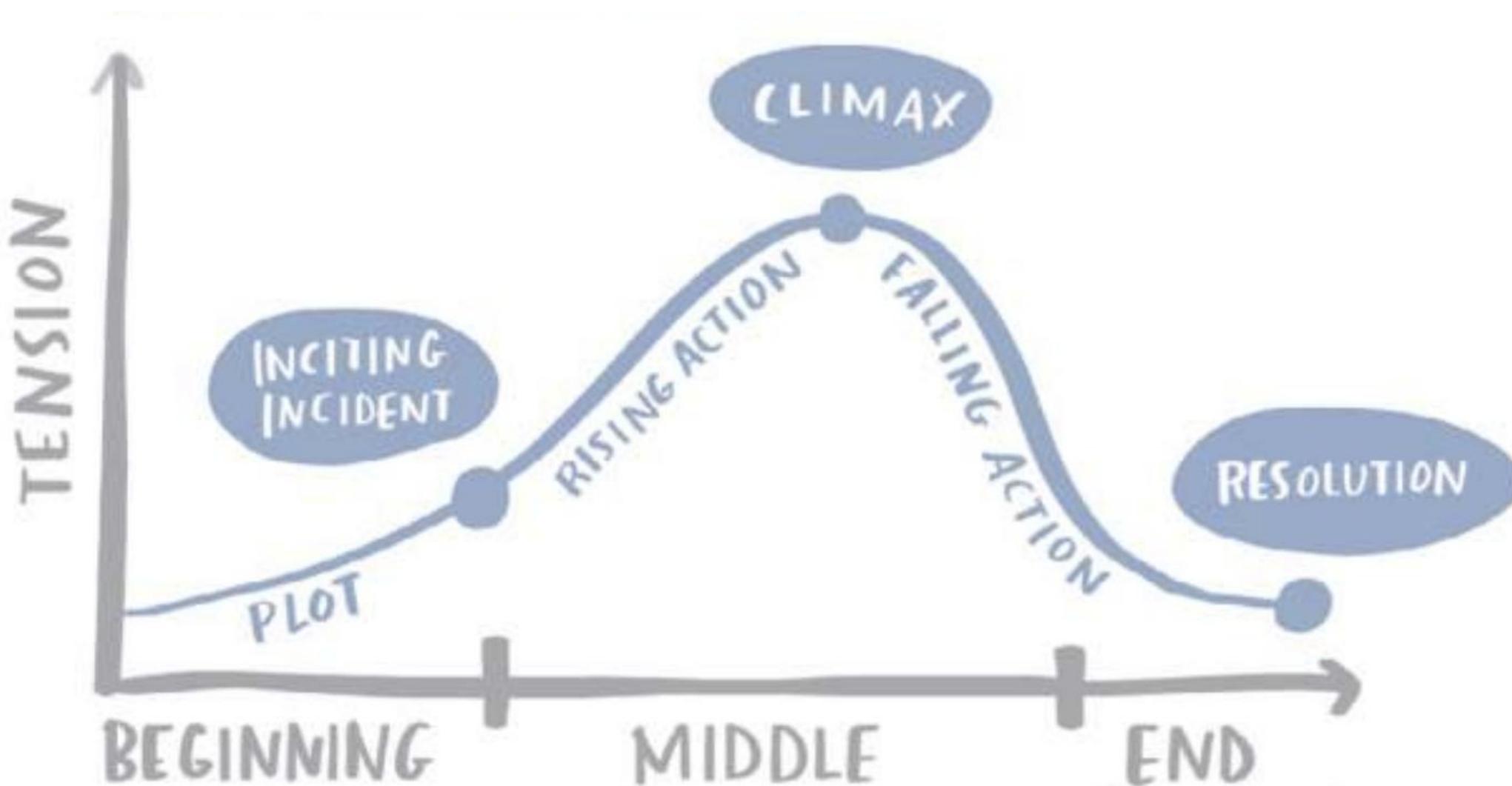
- ▶ Large parts of the population are faced with unbearable injustice
- ▶ What can be done to overcome it?

5. Quest for justice



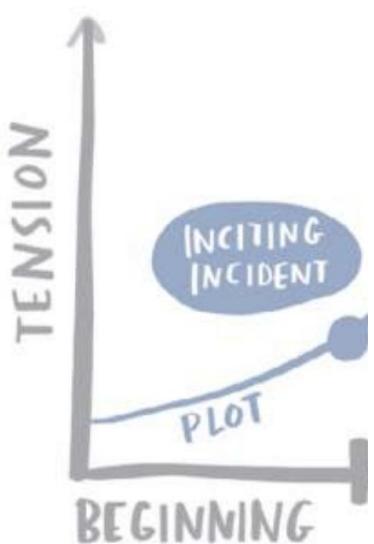


Narrative Arc



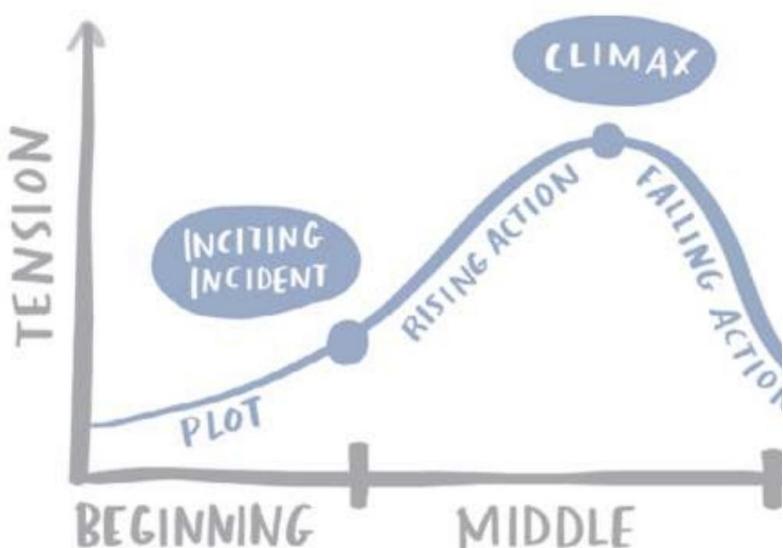
Beginning

- ▶ **Setting:** when and where does the story take place?
- ▶ **Main character:** who is driving the action?
- ▶ **Imbalance:** why is it necessary, what has changed?



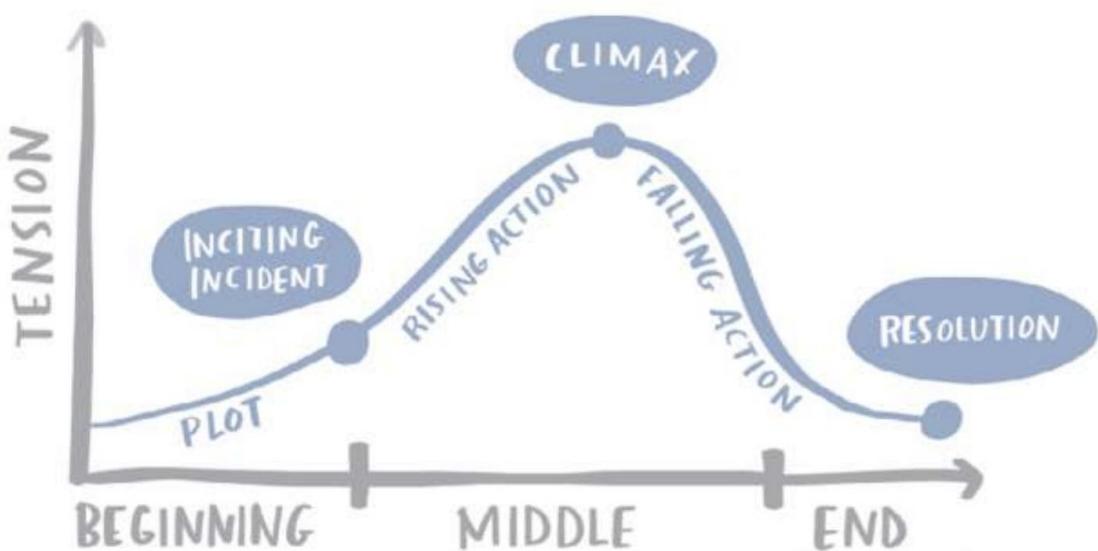
Middle

- ▶ Background information
- ▶ External comparison points
- ▶ Examples that illustrate the issue
- ▶ Show what happens if no action is taken
- ▶ Reveal actions that can help solve the problem

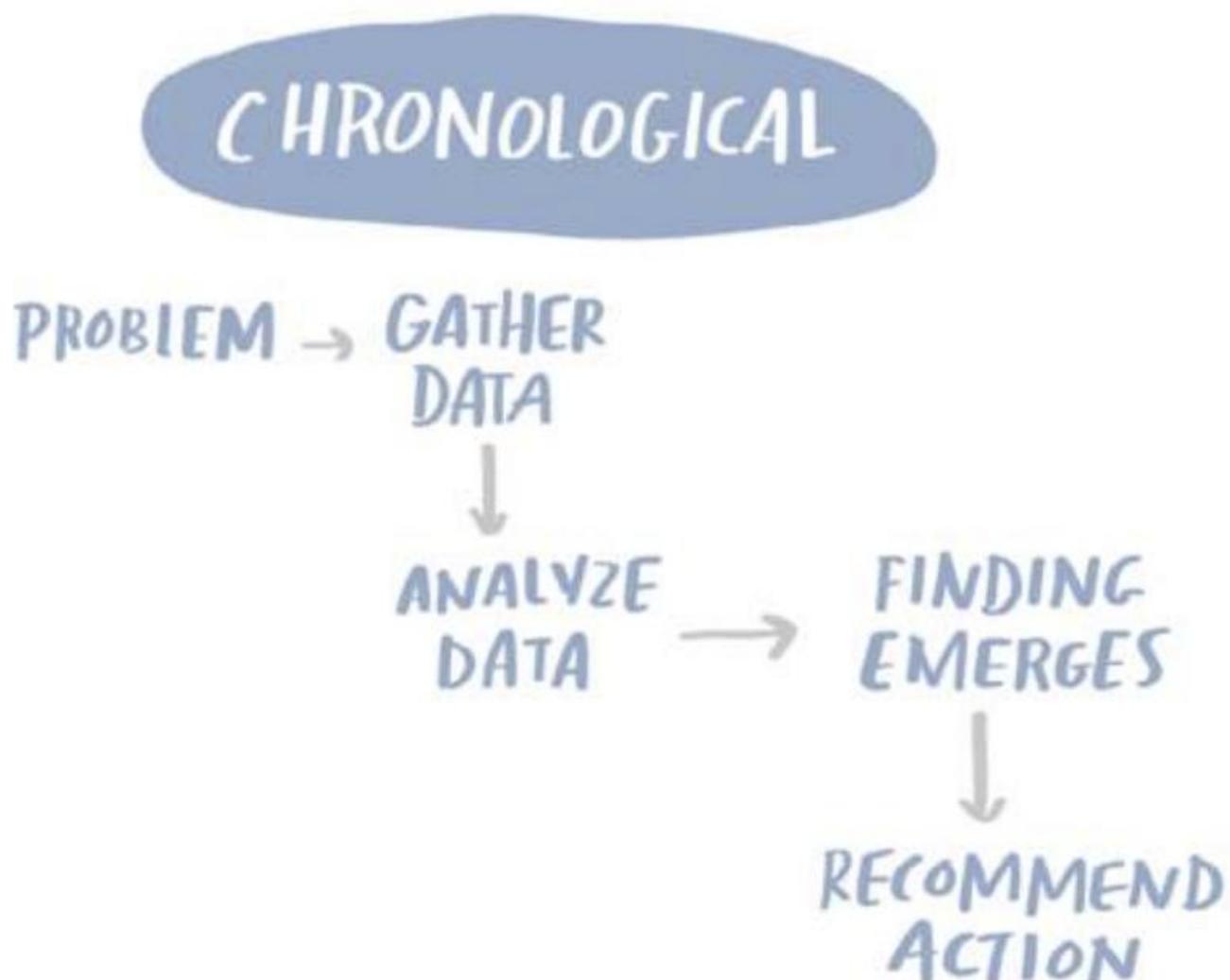


End

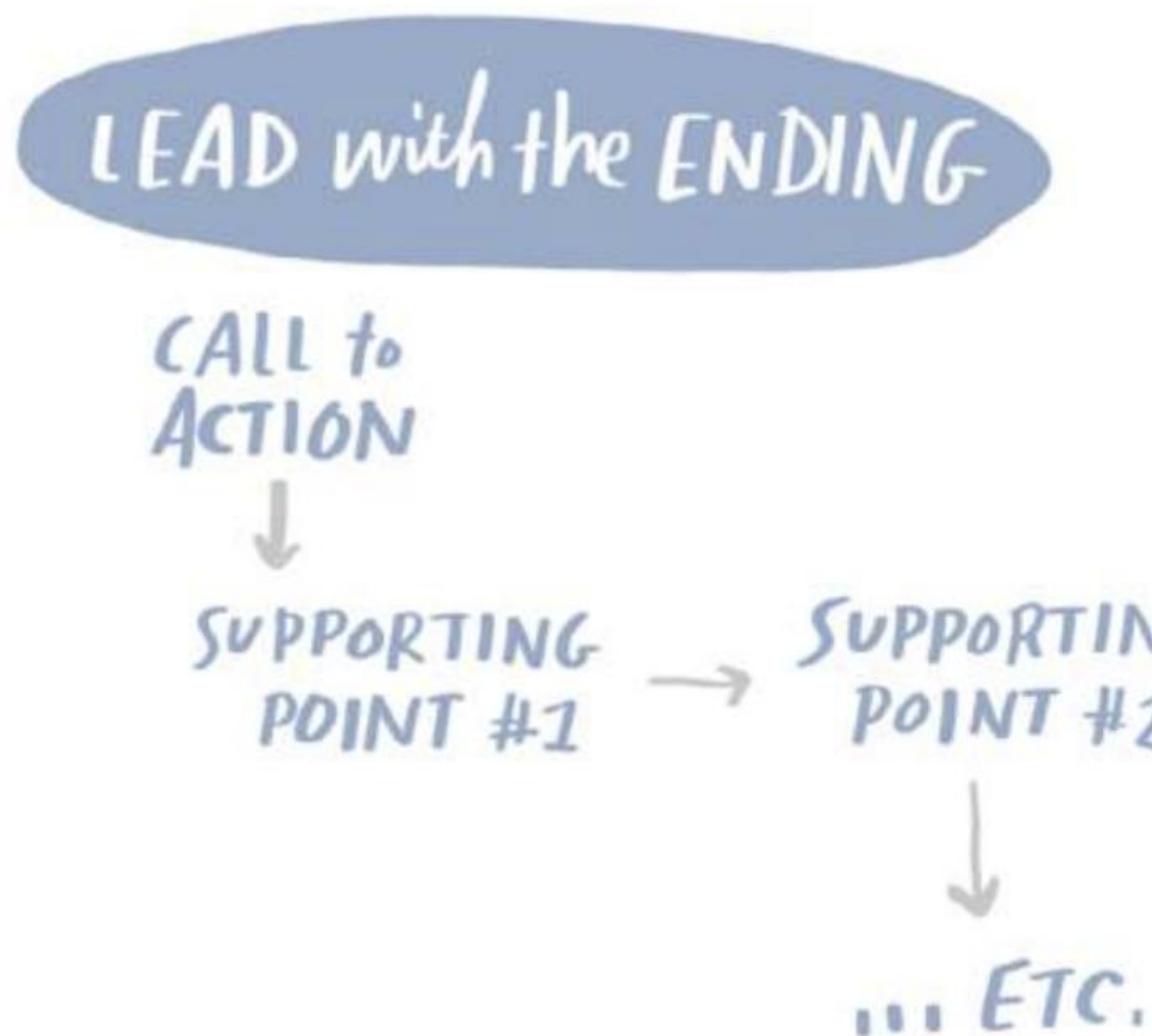
If possible, end with a call for action



Chronological flow



Lead-with-ending flow



Chronological vs. lead-with-ending

The story should have a logical order, but which one may depend on the setting:

- ▶ Do you first need to establish credibility? → *Chronological*
- ▶ Busy / impatient audience? → *Lead-with-ending*
- ▶ Do they care about your process (→ *chronological*) or just want the answer (→ *lead-with-ending*)?



Repetition and Reinforcement

The Power of Repetition

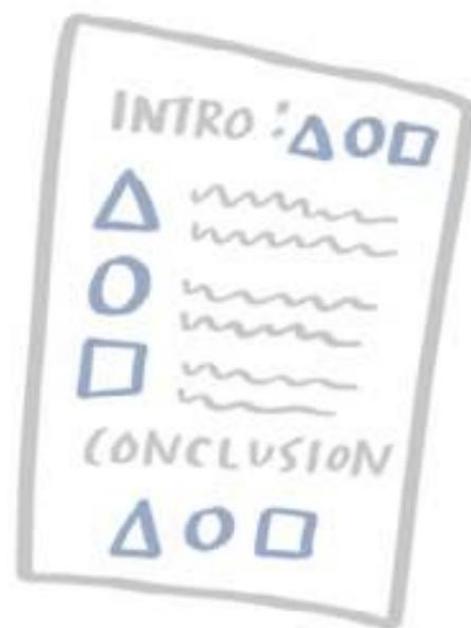
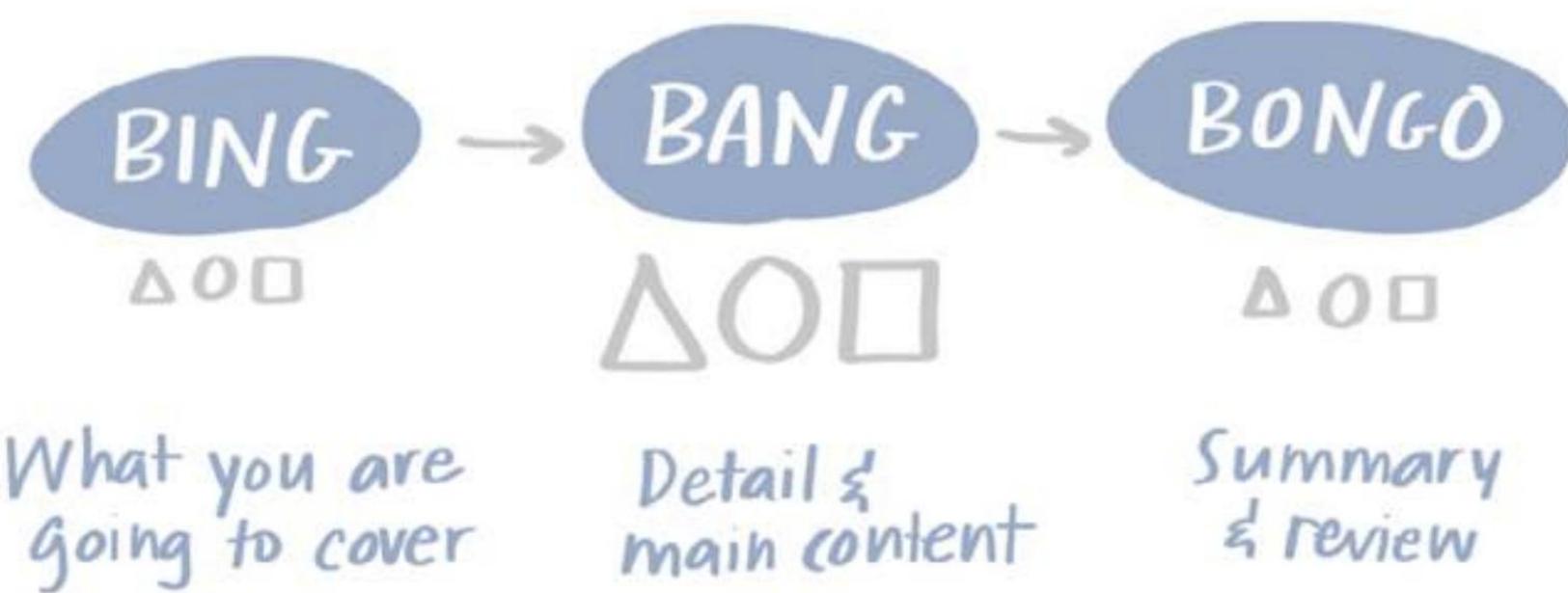
Short-term to long-term memory

- ▶ Repetition helps transferring information to long-term memory

Isn't this boring?

- ▶ No, it may feel redundant for you, because you are familiar with the content, but for your audience your contents are likely new
- ▶ Reinforcement (provide examples, reference points, complementary information...) helps strengthening the message, and making it more varied

The Power of Repetition



Bing In a Scientific Publication

Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings

Tolga Bolukbasi¹, Kai-Wei Chang², James Zou², Venkatesh Saligrama^{1,2}, Adam Kalai²

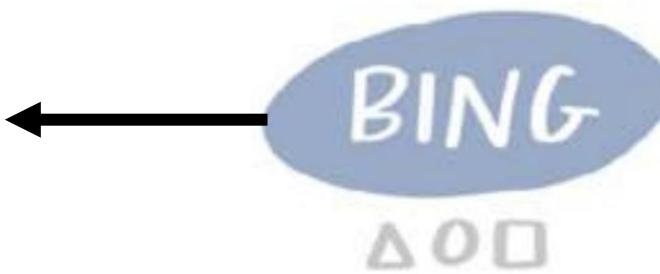
¹Boston University, 8 Saint Mary's Street, Boston, MA

²Microsoft Research New England, 1 Memorial Drive, Cambridge, MA

tolgab@bu.edu, kw@kwchang.net, jamesyzou@gmail.com, srv@bu.edu, adam.kalai@microsoft.com

Abstract

The blind application of machine learning runs the risk of amplifying biases present in data. Such a danger is facing us with *word embedding*, a popular framework to represent text data as vectors which has been used in many machine learning and natural language processing tasks. We show that even word embeddings trained on Google News articles exhibit female/male gender stereotypes to a disturbing extent. This raises concerns because their widespread use, as we describe, often tends to amplify these biases. Geometrically, gender bias is first shown to be captured by a direction in the word embedding. Second, gender neutral words are shown to be linearly separable from gender definition words in the word embedding. Using these properties, we provide a methodology for modifying an embedding to remove gender stereotypes, such as the association between the words *receptionist* and *female*, while maintaining desired associations such as between the words *queen* and *female*. Using crowd-worker evaluation as well as standard benchmarks, we empirically demonstrate that our algorithms significantly reduce gender bias in embeddings while preserving its useful properties such as the ability to cluster related concepts and to solve analogy tasks. The resulting embeddings can be used in applications without amplifying gender bias.



Bang

In a Scientific Publication

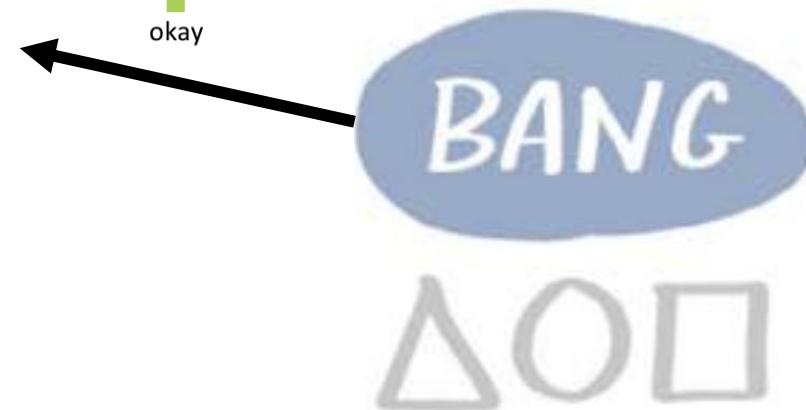
3 Geometry of Gender and Bias in Word Embeddings

Extreme <i>she</i>	Extreme <i>he</i>
1. homemaker	1. maestro
2. nurse	2. skipper
3. receptionist	3. protege
4. librarian	4. philosopher
5. socialite	5. captain
6. hairdresser	6. architect
7. nanny	7. financier
8. bookkeeper	8. warrior
9. stylist	9. broadcaster
10. housekeeper	10. magician



4 Debiasing algorithms

6 Debiasing results



Bongo

In a Scientific Publication

7 Discussion

Word embeddings help us further our understanding of bias in language. We find a single direction that largely captures gender, that helps us capture associations between gender neutral words and gender as well as indirect inequality. The projection of gender neutral words on this direction enables us to quantify their degree of female- or male-bias.

