# Growing Polarity & Sensationalization in the News

Jane Shclover

BrainStation, Data Science Diploma
Capstone Project
August 2022

# Context

# Has polarity in headline language escalated over time?

**Sources**

The New York Times

NYT Headlines
1981-2021

**Methods**

Natural Language Processing
vaderSentiment

**Tools**

Python
Tableau

**Additional Summary & Code**

github.com/janeshclover/nyt-nlp

# Method

# Data Preparation

**1**

## Initial State

/extract from NYT API
by month, 1981-2021

**Shape**
3.9m rows
28 columns

**2**

## Clean & Prep

/remove empty data
/clean headline data
- Erroneous values
- Duplicate values
- Non-English values
/filter to top 15 news categories
/remove all columns except
headline, category, dates

**Shape**
2.7m rows
4 columns

**3**

## Polarity Assignment

/add polarity scores
- Neutral
- Negative
- Positive
- Compound

**Shape**
2.7m rows
8 columns

**4**

## Tokenization

/tokenize highest and lowest
1k of headlines for every 5
years of data, only within
`World` news section

**Shape**
1.0k rows
1.5k-3.0k columns

`VaderSentiment` assigns a series of sentiment scores to text. In this case, based on its contents, each headline receives one of the following:

**Positive Score**

0                                          1
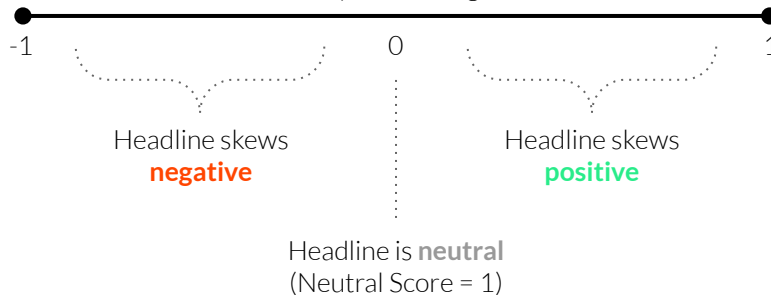
**Negative Score**

0                                          1

**Neutral Score**

0                                          1

**Compound Score**
(combination of positive, negative, neutral)

-1                          0                          1

Headline skews
**negative**

Headline skews
**positive**

Headline is **neutral**
(Neutral Score = 1)

# How does this look in practice?

vaderSentiment assigns scores based on presence of negative, positive, and neutral sentiment in a given string of text [headline].

## Sample Headlines Mapped Against Positive & Negative Scores

Negative Score

**Homegrown Terrorism**
January 2018
*Opinion*

Negative: 0.821
Positive: 0.0
Neutral: 0.179
Compound: -.683

*No articles with both negative and positive score > than .5*

Positive Score

**November New Home Sales Up**
December 1981
*Business Day*

Negative: 0
Positive: 0
Neutral: 1
Compound: 0

**Gore Shares Nobel Peace Prize**
October 2007
*World*

Negative: 0.0
Positive: 0.818
Neutral: 0.182
Compound: .840

# How does this look in practice?

vaderSentiment assigns scores based on presence of negative, positive, and neutral sentiment in a given string of text [headline].
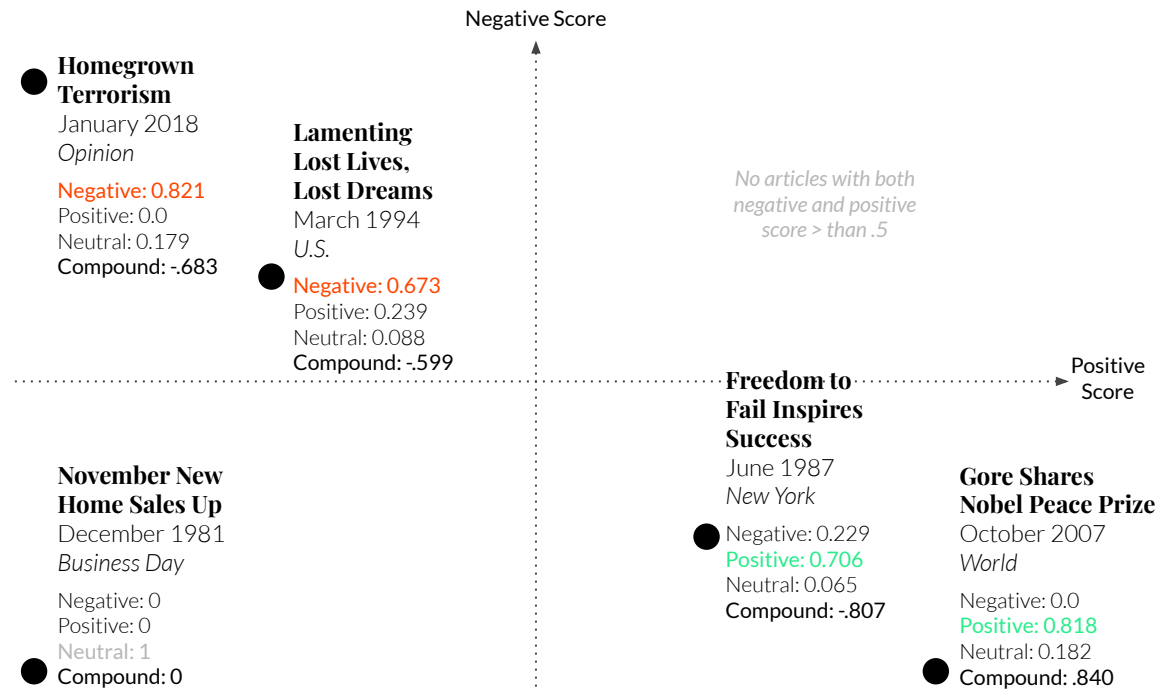
## Sample Headlines Mapped Against Positive & Negative Scores

Negative Score

**Homegrown Terrorism**
January 2018
*Opinion*

Negative: 0.821
Positive: 0.0
Neutral: 0.179
**Compound: -.683**

**Lamenting Lost Lives, Lost Dreams**
March 1994
*U.S.*

Negative: 0.673
Positive: 0.239
Neutral: 0.088
**Compound: -.599**

*No articles with both negative and positive score > than .5*

Positive Score

**Freedom to Fail Inspires Success**
June 1987
*New York*

Negative: 0.229
Positive: 0.706
Neutral: 0.065
**Compound: -.807**

**Gore Shares Nobel Peace Prize**
October 2007
*World*

Negative: 0.0
Positive: 0.818
Neutral: 0.182
**Compound: .840**

**November New Home Sales Up**
December 1981
*Business Day*

Negative: 0
Positive: 0
Neutral: 1
**Compound: 0**

# How does this look in practice?

vaderSentiment assigns scores based on presence of negative, positive, and neutral sentiment in a given string of text [headline].
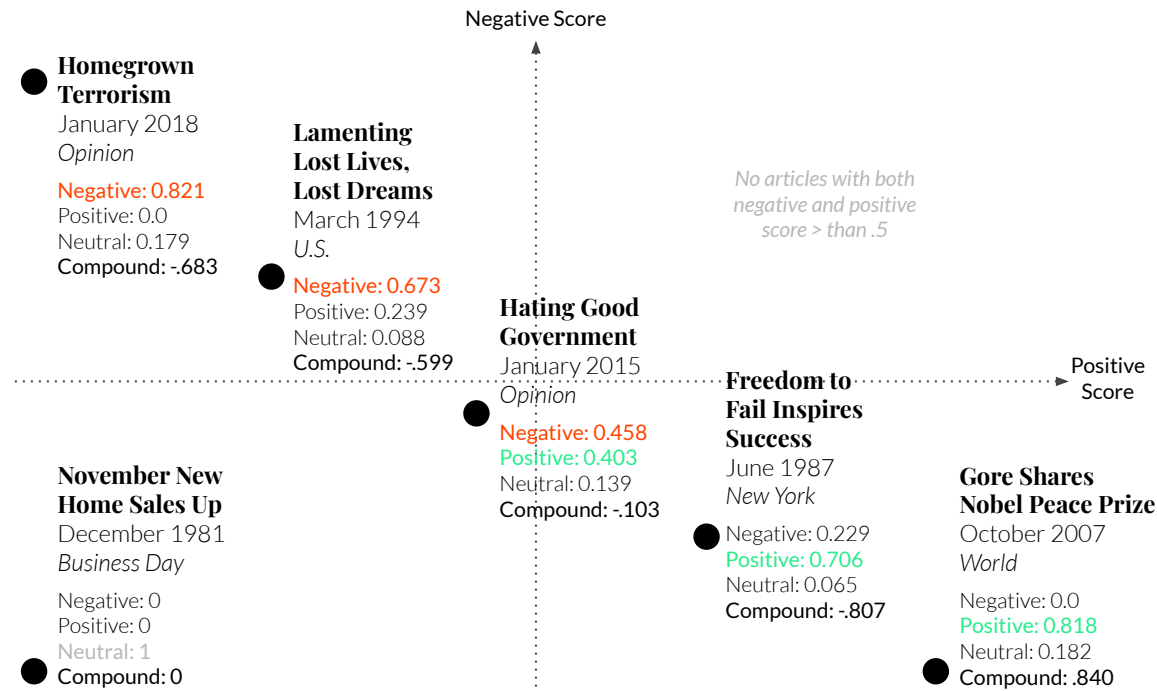
## Sample Headlines Mapped Against Positive & Negative Scores

Negative Score

**Homegrown Terrorism**
January 2018
*Opinion*

Negative: 0.821
Positive: 0.0
Neutral: 0.179
**Compound: -.683**

**Lamenting Lost Lives, Lost Dreams**
March 1994
*U.S.*

Negative: 0.673
Positive: 0.239
Neutral: 0.088
**Compound: -.599**

*No articles with both negative and positive score > than .5*

**Hating Good Government**
January 2015
*Opinion*

Negative: 0.458
Positive: 0.403
Neutral: 0.139
**Compound: -.103**

Positive Score

**Freedom to Fail Inspires Success**
June 1987
*New York*

Negative: 0.229
Positive: 0.706
Neutral: 0.065
**Compound: -.807**

**November New Home Sales Up**
December 1981
*Business Day*

Negative: 0
Positive: 0
Neutral: 1
**Compound: 0**

**Gore Shares Nobel Peace Prize**
October 2007
*World*

Negative: 0.0
Positive: 0.818
Neutral: 0.182
**Compound: .840**

Source: 1981-2021 headline data exported from NYT API | Method: vaderSentiment applied in Python polarity scores

# Analysis

# Over the past 40 years, there has been a decline in the share of neutral news

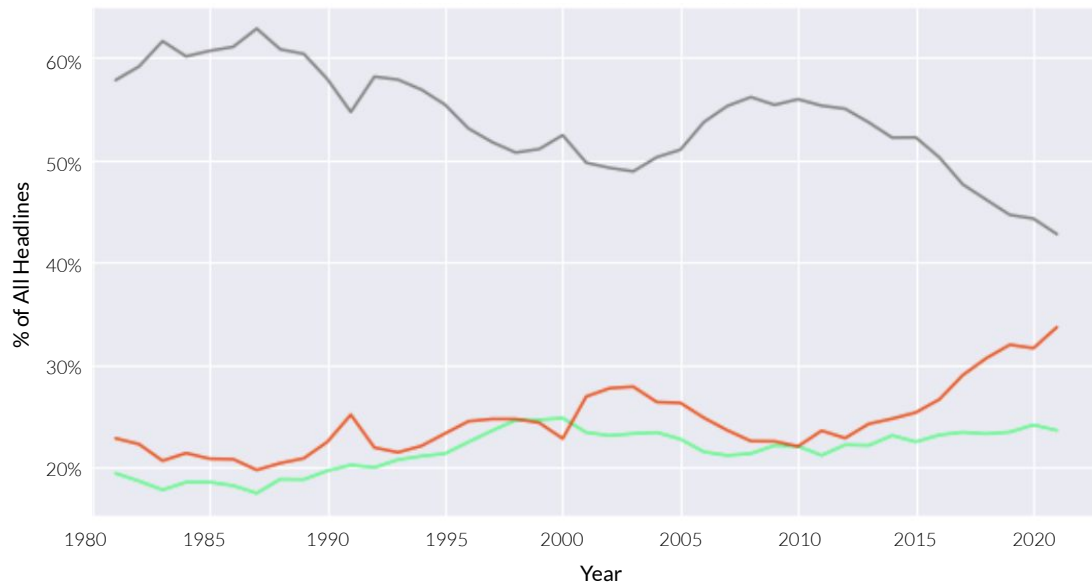Since 1981, the share of net negative and positive headlines has been on the rise.

## Headline Net Sentiment Share 1981-2021

● Net Neutral Headlines
*Compound score = 0*

● Net Negative Headlines
*Compound score < 0*

● Net Positive Headlines
*Compound score > 0*



Source: 1981-2021 headline data exported from NYT API | Method: vaderSentiment applied in Python polarity scores

# Headlines have consistently migrated from net neutrality towards more charged language

The share of net neutral headlines has declined each decade since the 80s, currently at its lowest in the early 2020s, when less than ½ of headlines so far are considered neutral.

## Compound Score Distribution by Decade

*Compound scores range from -1 to 1, from most negative to most positive.*
*A compound score of 0 signifies complete neutrality.*
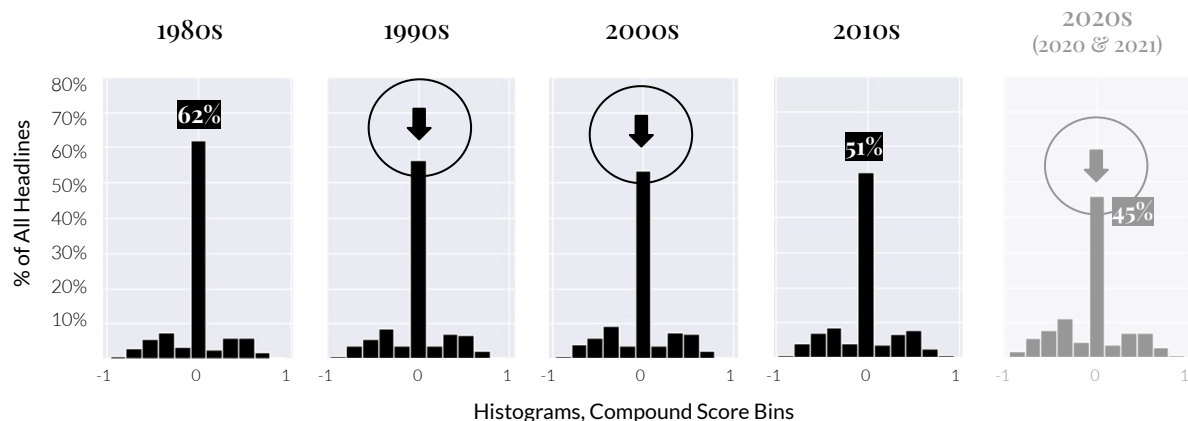
# The shift is not one-dimensional, but marked by a growing *intensity* in sentiment

Distribution of positive scores has gradually migrated upwards with each passing decade, signifying increasing positivity in headline language.

## Positive Score Distribution by Decade

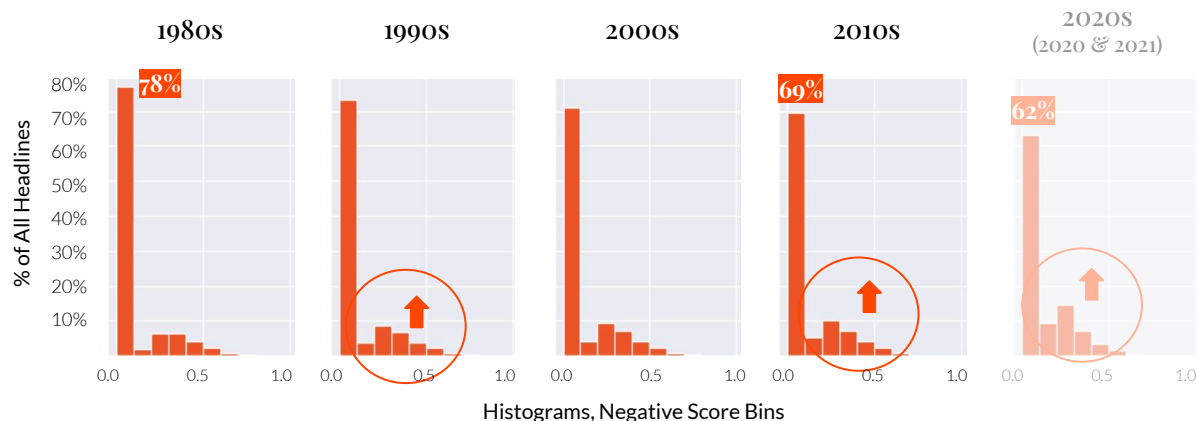*Positive scores range from 0 to 1, from least to most negative.*



Histograms, Positive Score Bins

% of All Headlines

1980s · 1990s · 2000s · 2010s · 2020s (2020 & 2021)

# The stronger trade-off has been towards negativity

The distribution of negative scores has crept higher up the chart every decade, moving at an accelerated pace compared to positive scores*.

*Positive scores saw a migration of **6pp** from the first to higher bins between 1980s and 2010s, vs. **9pp** for Negative score migration.

## Negative Score Distribution by Decade

*Negative scores range from 0 to 1, from least to most negative.*



% of All Headlines

1980s   1990s   2000s   2010s   2020s (2020 & 2021)

78%   69%   62%

Histograms, Negative Score Bins

# The same trend is echoed across top categories of news

`World`, `New York`, and `Business Day` stand out as sections where the exchange has been most extreme.
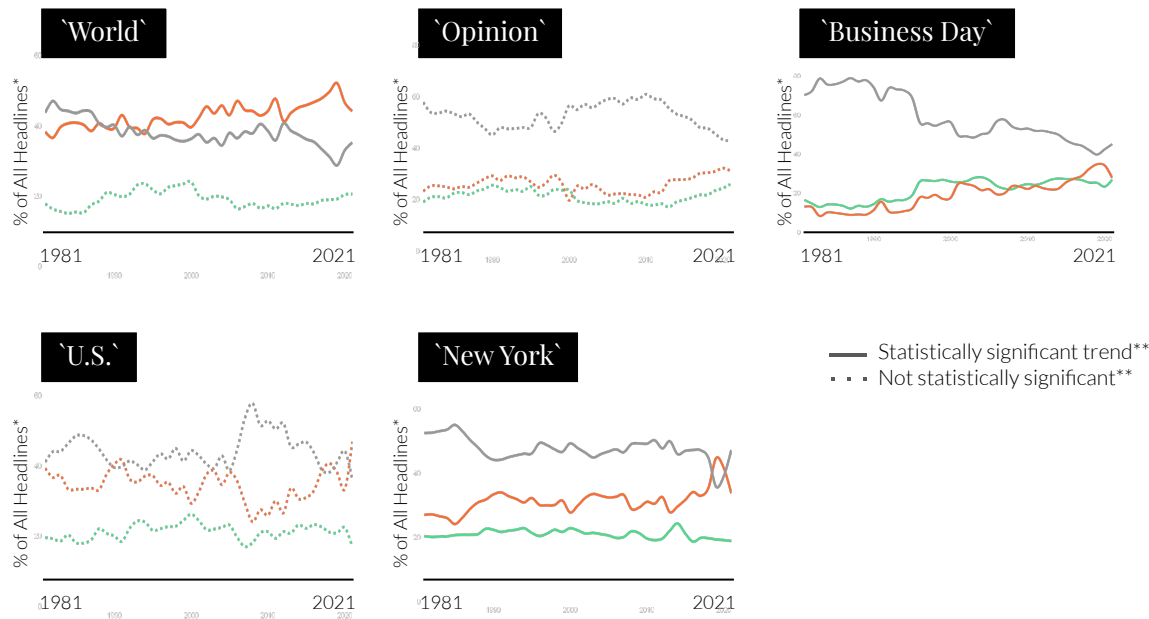
`World` equally stands out for its share of negative headlines exceeding all others since the 1990s.

## Net Sentiment Trends by News Category

● Net Neutral Headlines
*Compound score = 0*
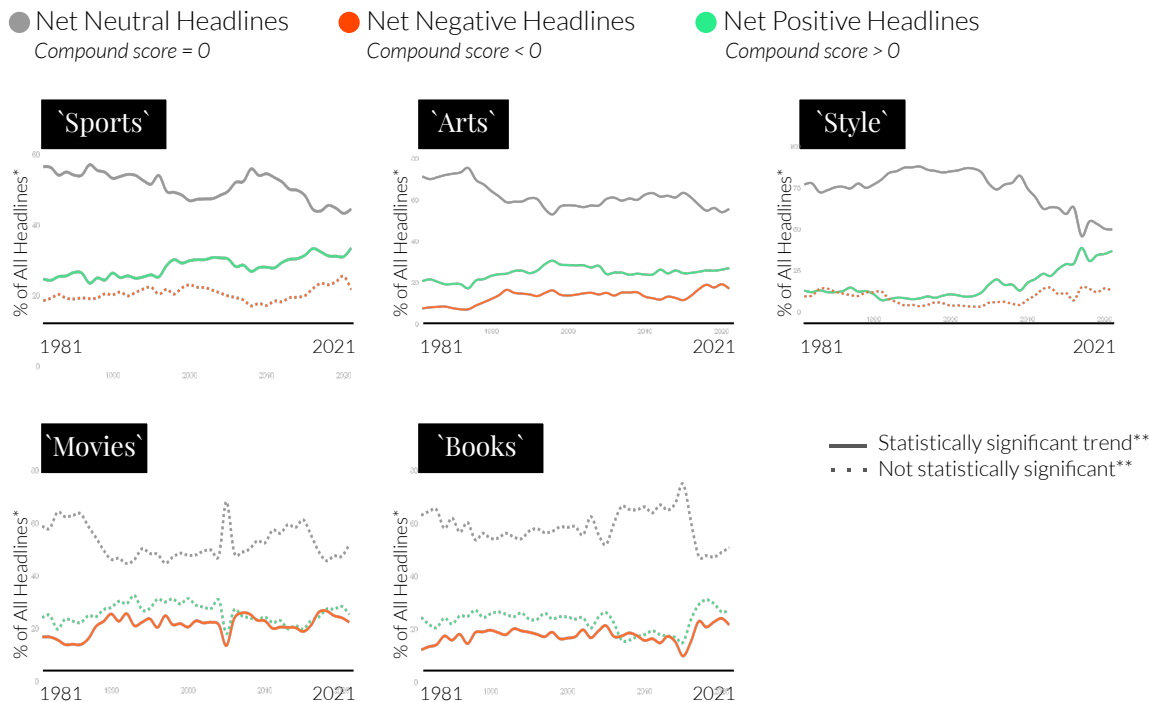
● Net Negative Headlines
*Compound score < 0*

● Net Positive Headlines
*Compound score > 0*



`World`

`Opinion`

`Business Day`

`U.S.`

`New York`

—— Statistically significant trend**
⋯⋯ Not statistically significant**

Even some lifestyle categories have been subject to this shift, though not as dramatically

`Sports`, `Arts`, and `Style` content have become increasingly 'emotional' over the past 40 years.

## Net Sentiment Trends by News Category

● Net Neutral Headlines
*Compound score = 0*

● Net Negative Headlines
*Compound score < 0*

● Net Positive Headlines
*Compound score > 0*



`Sports`

% of All Headlines*

1981 — 2021



`Arts`

% of All Headlines*

1981 — 2021



`Style`

% of All Headlines*

1981 — 2021



`Movies`

% of All Headlines*

1981 — 2021



`Books`

% of All Headlines*

1981 — 2021

—— Statistically significant trend**
······ Not statistically significant**

*Axis lines not synchronized across category charts. **Statistical significance based on p value of < .05
Source: 1981-2021 headline data exported from NYT API | Method: vaderSentiment applied in Python polarity scores
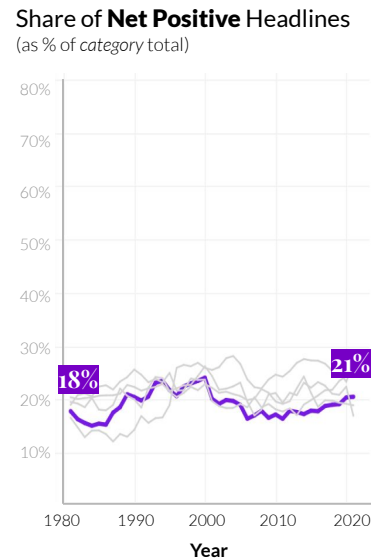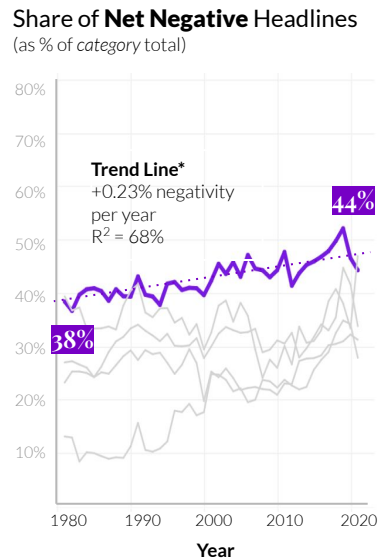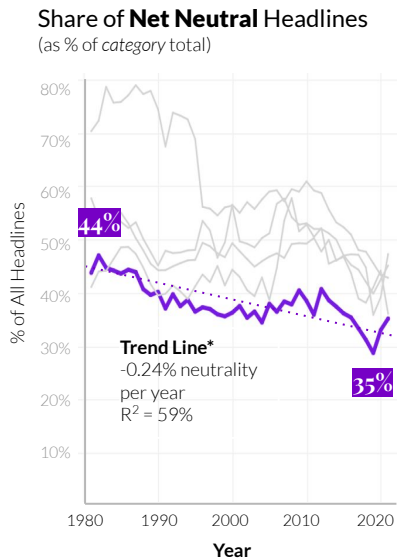
# Historically, `World` has been the most grim section of the news as compared to other categories.

This category has consistently trended towards more negative news, with 44% of published headlines communicating negativity in 2021, and peaking at over 50% in 2018.

## Comparing Net Sentiment Trends Across Categories

● World  ● Other categories

### Share of **Net Neutral** Headlines
(as % of *category* total)

**Trend Line***
-0.24% neutrality per year
R² = 59%

44%

35%

% of All Headlines

Year

### Share of **Net Negative** Headlines
(as % of *category* total)

**Trend Line***
+0.23% negativity per year
R² = 68%

44%

38%

Year

### Share of **Net Positive** Headlines
(as % of *category* total)

21%

18%

Year

# Scorecard: `World`

/Volume of headlines in this category have fluctuated over time, with a majority of net negative & net neutral
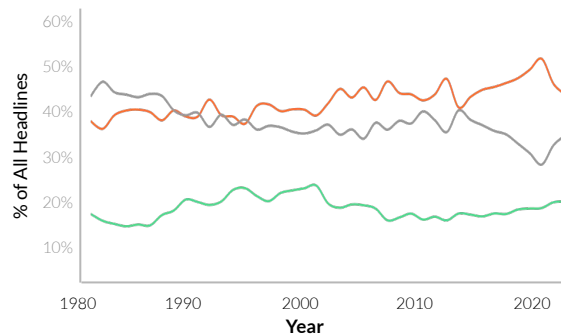
/Net negative headlines overshadow others beginning in early 90s

/Neutrality is relatively low for this category to start with, but drops even lower by 2010s Between 1980s and 2010s
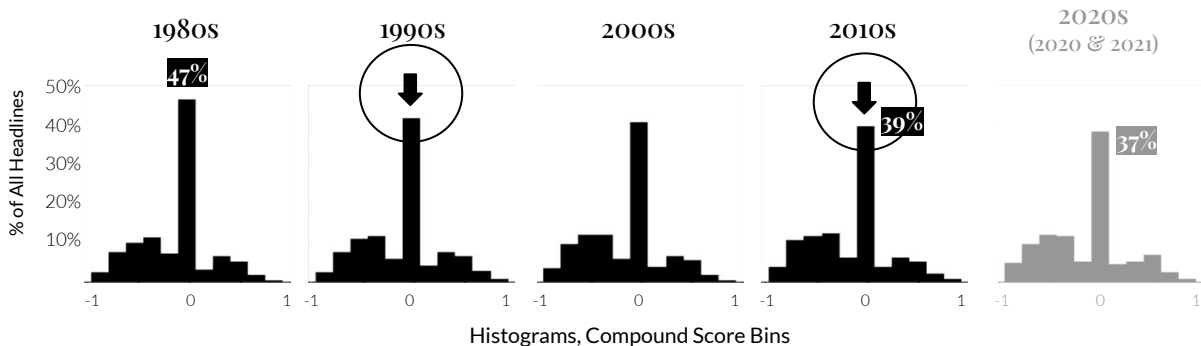
## Net Sentiment Trend, Volume
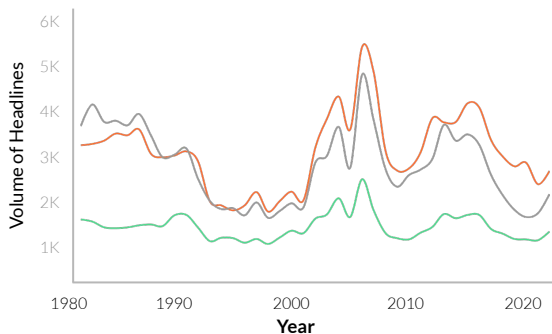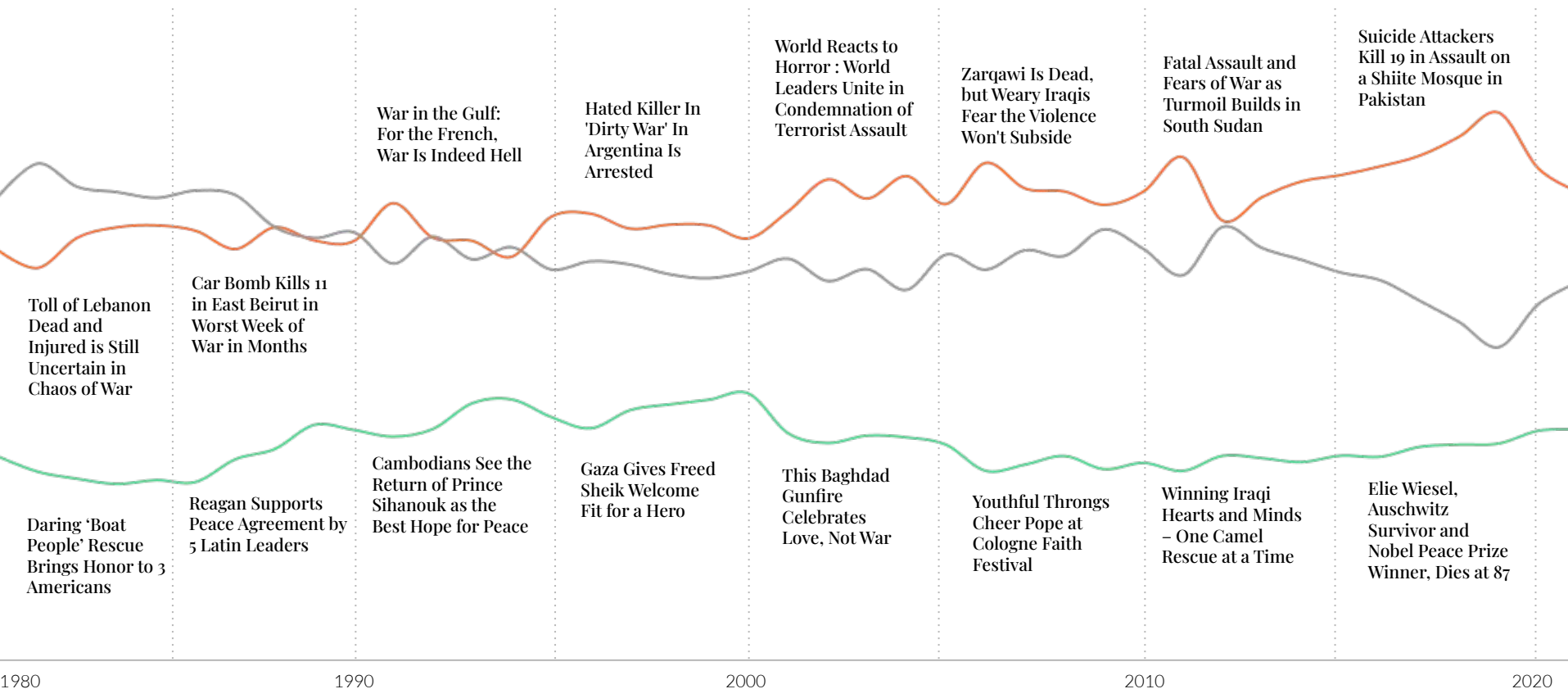
## Net Sentiment Trend, % of Total
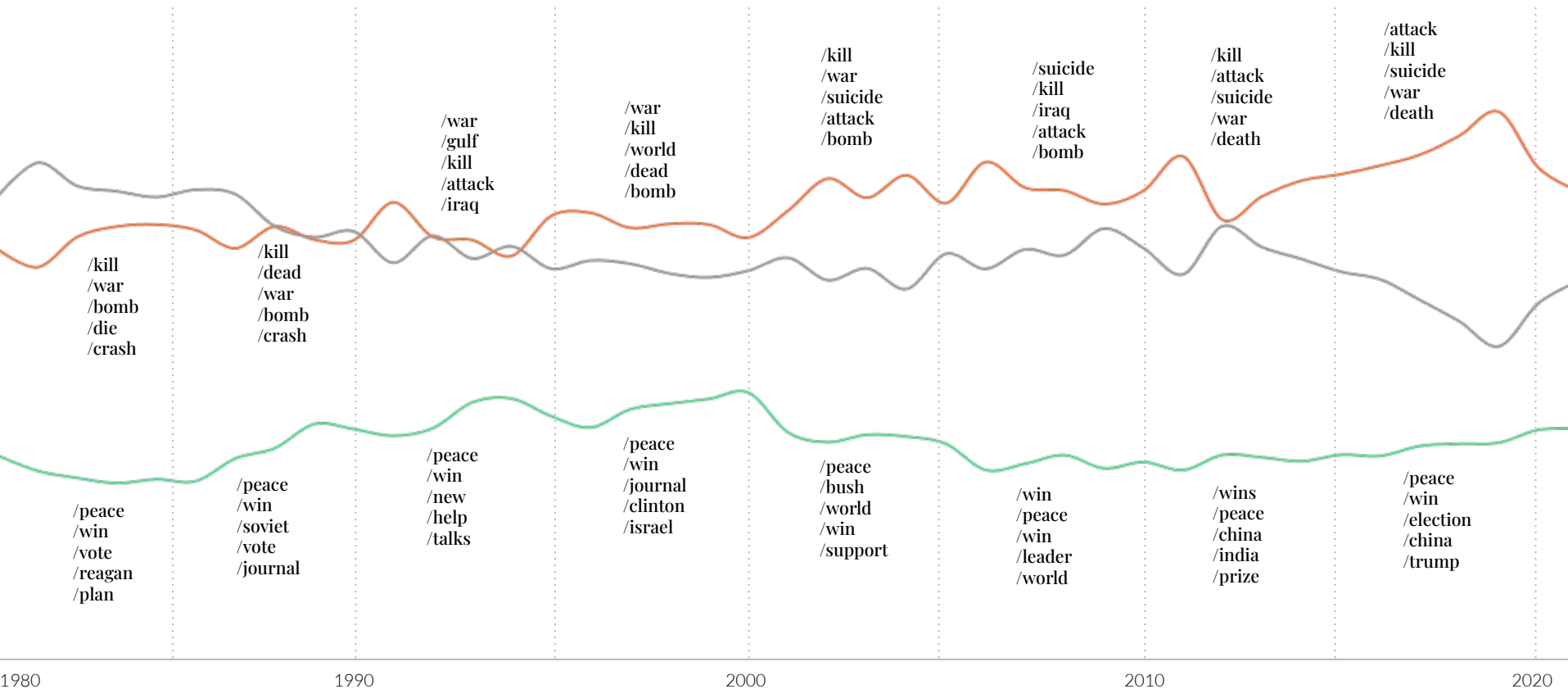
## Compound Score Distribution by Decade

### 1980s

### 1990s

### 2000s

### 2010s

### 2020s
(2020 & 2021)

Histograms, Compound Score Bins

# `World` Section Headline Samples

by lowest and highest compound scores, every 5 years



World Reacts to Horror : World Leaders Unite in Condemnation of Terrorist Assault

War in the Gulf: For the French, War Is Indeed Hell

Hated Killer In 'Dirty War' In Argentina Is Arrested

Zarqawi Is Dead, but Weary Iraqis Fear the Violence Won't Subside

Fatal Assault and Fears of War as Turmoil Builds in South Sudan

Suicide Attackers Kill 19 in Assault on a Shiite Mosque in Pakistan

Toll of Lebanon Dead and Injured is Still Uncertain in Chaos of War

Car Bomb Kills 11 in East Beirut in Worst Week of War in Months

Daring 'Boat People' Rescue Brings Honor to 3 Americans

Reagan Supports Peace Agreement by 5 Latin Leaders

Cambodians See the Return of Prince Sihanouk as the Best Hope for Peace

Gaza Gives Freed Sheik Welcome Fit for a Hero

This Baghdad Gunfire Celebrates Love, Not War

Youthful Throngs Cheer Pope at Cologne Faith Festival

Winning Iraqi Hearts and Minds – One Camel Rescue at a Time

Elie Wiesel, Auschwitz Survivor and Nobel Peace Prize Winner, Dies at 87

1980          1990          2000          2010          2020

# `World` Section Token Samples

by frequency amongst headlines with lowest and highest compound scores, every 5 years

/attack
/kill
/suicide
/war
/death

/kill
/war
/suicide
/attack
/bomb

/suicide
/kill
/iraq
/attack
/bomb

/kill
/attack
/suicide
/war
/death

/war
/gulf
/kill
/attack
/iraq

/war
/kill
/world
/dead
/bomb

/kill
/war
/bomb
/die
/crash

/kill
/dead
/war
/bomb
/crash

/peace
/win
/new
/help
/talks

/peace
/win
/journal
/clinton
/israel

/peace
/bush
/world
/win
/support

/wins
/peace
/china
/india
/prize

/peace
/win
/election
/china
/trump

/peace
/win
/vote
/reagan
/plan

/peace
/win
/soviet
/vote
/journal

/win
/peace
/win
/leader
/world

1980          1990          2000          2010          2020

# Conclusion

**Conclusion**

There has been a migration from mild to increasingly charged headline language in many categories of the NYT's news. In these instances **neutrality** has slowly been replaced with both **positive**, but more often, **negative**, sentiment.

**1**

## Deeper language & topic evolution exploration

Look deeper into how headline language may change over time (how is the same topic discussed today vs. 10, 20, 40 years ago?)

**2**

## News category scorecards

Creating & comparing a scorecard for each category of news that speaks to its evolving sentiment and the content inputs driving this

**3**

## Explore other variables related to sentiment scores

Explore the relationship between things like headline length with assigned polarity scores. Test other sentiment tools beyond vader!

**4**

## Deeper investigation into righteousness & sensationalization

Version 2.0 of this project is digging deeper into the actual quality and character of language to validate to bring suggestion about sensationalization home.