

# Автоматический анализ тематики текста

от частотных слов — к вероятностному тематическому  
моделированию

Часть II: Тематическое моделирование

Даня Скоринкин, DH HSE

30 апреля 2020

# Тематическое моделирование в Digital Humanities: кейсы



# Mining the Dispatch

# DAILY DISPATCH.

VOLUME XXIX

RICHMOND, VA., SATURDAY, APRIL 4, 1865.

ONE HUNDRED AND TWENTY-FIVE CENTS

# FLAMES!

**WATERFRONT ABLAZE AS UNION ARRIVES**

*Citizens struggle to evacuate as fire ravages city*

Union General  
Weitzel takes com-  
mand of the city

CONFEDERATE OFFICIALS ABANDON  
RICHMOND IN THE NIGHT

Area between 7th  
and 15th streets a  
burned wasteland

Homes destroyed  
by the 100s

EDITOR'S NOTE: Owing to the tremen-  
dous loss that is the burning and  
destruction of our building, we here  
except more from the Richmond  
*Whig* than is normally reprinted.

THE CITY.

THE EVACUATION OF RICHMOND  
BY THE CONFEDERATE ARMY —  
THE OCCUPATION BY THE FED-



Минуточку... А какое нам дело до  
Гражданской войны в США?

# Минуточку... А какое нам дело до Гражданской войны в США?

1. Нам — никакого, но это одна из двух главных тем в американской истории

# Минуточку... А какое нам дело до Гражданской войны в США?

1. Нам — никакого, но это одна из двух главных тем в американской истории
2. Поэтому американские историки занимаются ей постоянно

# Минуточку... А какое нам дело до Гражданской войны в США?

1. Нам — никакого, но это одна из двух главных тем в американской истории
2. Поэтому американские историки занимаются ей постоянно
3. Американские университеты богатые и могут позволить себе больше хороших исследовательских проектов

# Минуточку... А какое нам дело до Гражданской войны в США?

1. Нам — никакого, но это одна из двух главных тем в американской истории
2. Поэтому американские историки занимаются ей постоянно
3. Американские университеты богатые и могут позволить себе больше хороших исследовательских проектов
4. Именно поэтому среди кейсов DH-исследований так много Гражданской войны в США и историй про 1770-е — 1790-е в США

# Richmond Daily Dispatch

- Газета Daily Dispatch издавалась в Ричмонде с 1850

# Richmond Daily Dispatch

- Газета Daily Dispatch издавалась в Ричмонде с 1850
- Ричмонд был столицей Конфедерации южных штатов почти всю Гражданскую войну

# Richmond Daily Dispatch

- Газета Daily Dispatch издавалась в Ричмонде с 1850
- Ричмонд был столицей Конфедерации южных штатов почти всю Гражданскую войну
- Газета запомнилась расизмом, топила за рабство, разжигала ненависть к северянам

# Richmond Daily Dispatch

- Газета Daily Dispatch издавалась в Ричмонде с 1850
- Ричмонд был столицей Конфедерации южных штатов почти всю Гражданскую войну
- Газета запомнилась расизмом, топила за рабство, разжигала ненависть к северянам
- Отступающие из Ричмонда конфедераты сожгли газету (вместе со всем городом) в начале апреля 1865



# Richmond Daily Dispatch

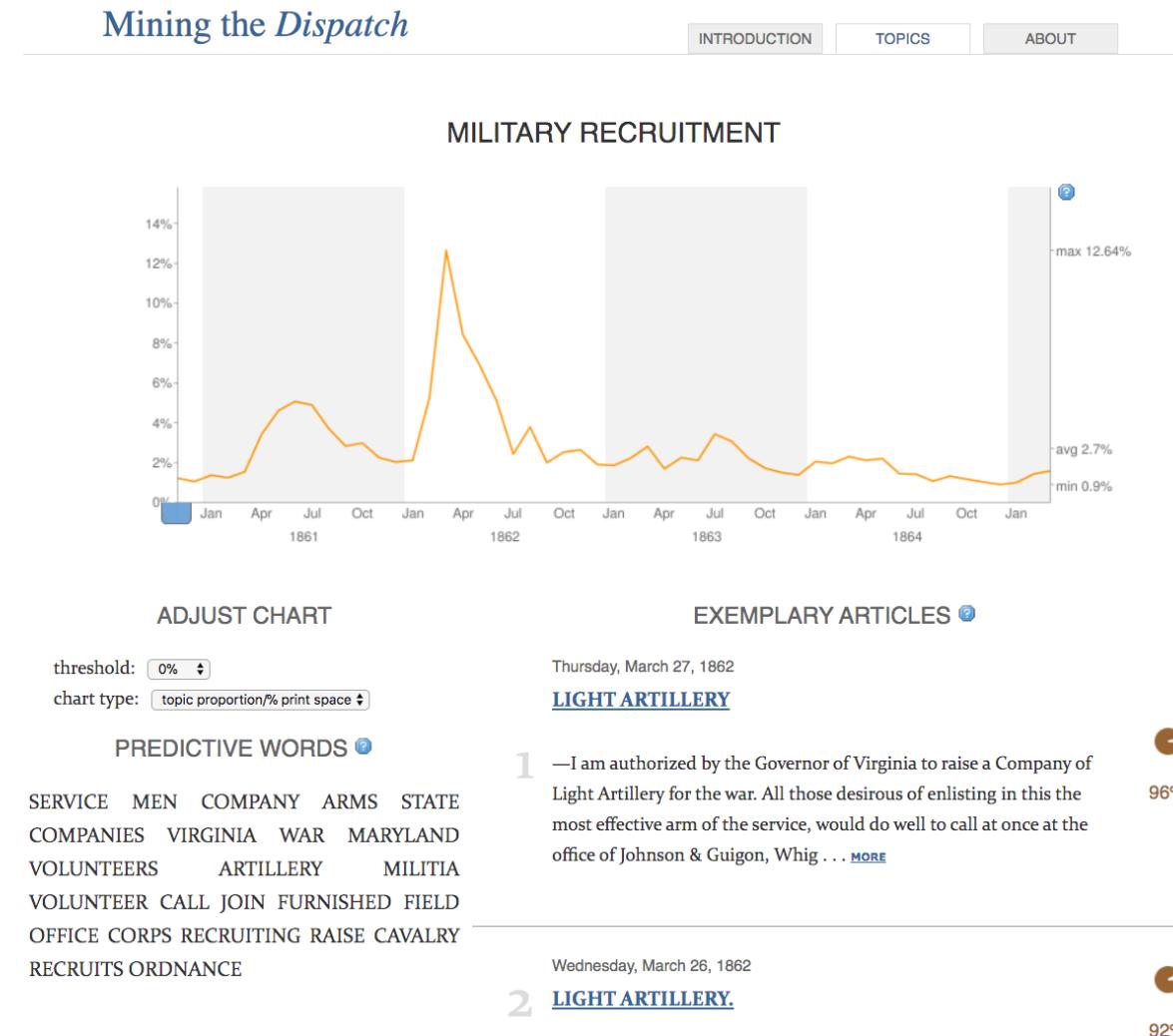
- Газета Daily Dispatch издавалась в Ричмонде с 1850
- Ричмонд был столицей Конфедерации южных штатов почти всю Гражданскую войну
- Газета запомнилась расизмом, топила за рабство, разжигала ненависть к северянам
- Отступающие из Ричмонда конфедераты сожгли газету (вместе со всем городом) в начале апреля 1865
- Сегодня Richmond Times-Dispatch
  - по-прежнему главная газета Ричмонда



# Проект Mining the Dispatch

- Тематическое моделирование архива Daily Dispatch
- Корпус: 112,000 выпусков или 24 млн. слов
- Метод: LDA
- Инструмент: MALLET ([mallet.cs.umass.edu](http://mallet.cs.umass.edu))
- Число топиков: 40
- Автор проекта: Robert K. Nelson, директор Digital Scholarship Lab в Университете Ричмонда
- Сайт: [dsl.richmond.edu/dispatch](http://dsl.richmond.edu/dispatch)

# На сайте можно исследовать темы в интерактивном режиме:



Что удалось выяснить с  
помощью тематического  
моделирования?

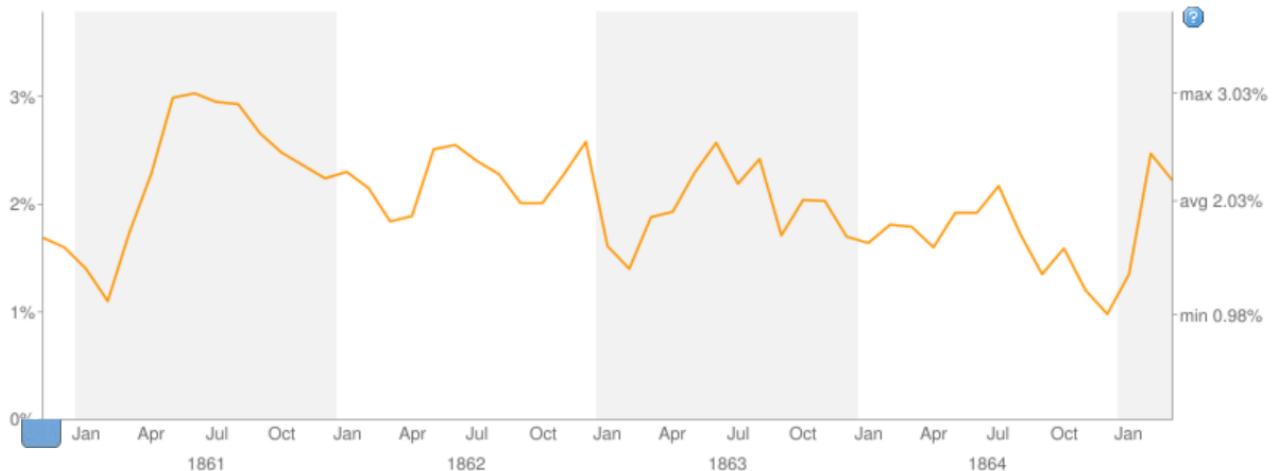
# Яростный национализм южан

«I will suggest that southerners (or at least the editor of the Dispatch) developed a particularly vitriolic version of Confederate nationalism to convince southern men to kill northerners. The Dispatch was full of articles that insisted that northerners were a foreign, unchristian, and uncivilized people wholly unlike southerners; in vicious editorials the Dispatch's editor maintained that northerners were infidels and beasts who it was not only okay but righteous to kill»

R. K. Nelson

# Тема «обличения северян»:

## ANTI-NORTHERN DIATRIBES



### ADJUST CHART

threshold:

chart type:

### PREDICTIVE WORDS

WAR PEOPLE SOUTH MEN SOUTHERN MAN  
NORTH WORLD YANKEE NORTHERN  
LINCOLN YANKEES COUNTRY BLOOD  
NATION HANDS MAKE TRUE HUMAN  
HISTORY POWER ENEMIES RACE HUMANITY

### EXEMPLARY ARTICLES

Monday, August 26, 1861

[Republican Tyranny.](#)

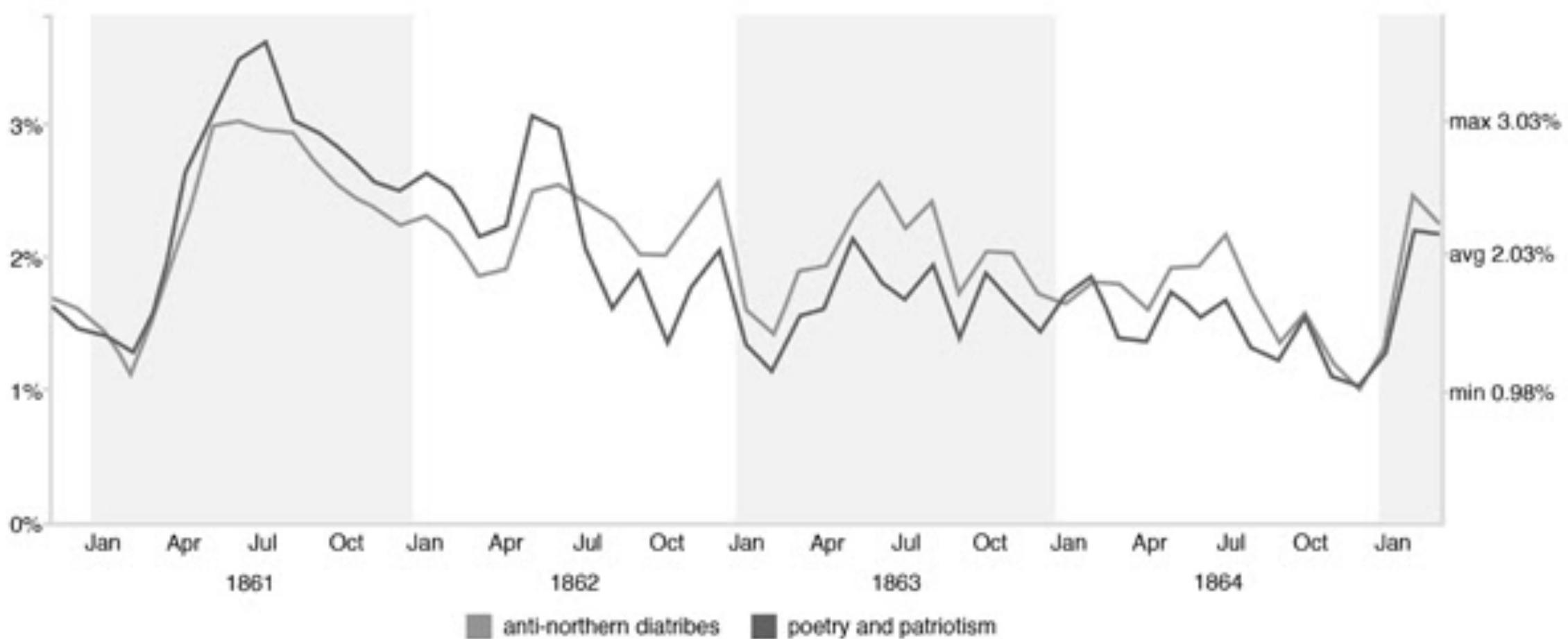
1

The brutal tyranny exercised by the Lincolniters in Maryland, Missouri, and wherever they have been able to obtain a foothold; the atrocities equalled only by those of the Sepays which their soldiery have committed, and the unheard of deed of despotism and darkness with which they threaten the . . . [MORE](#)

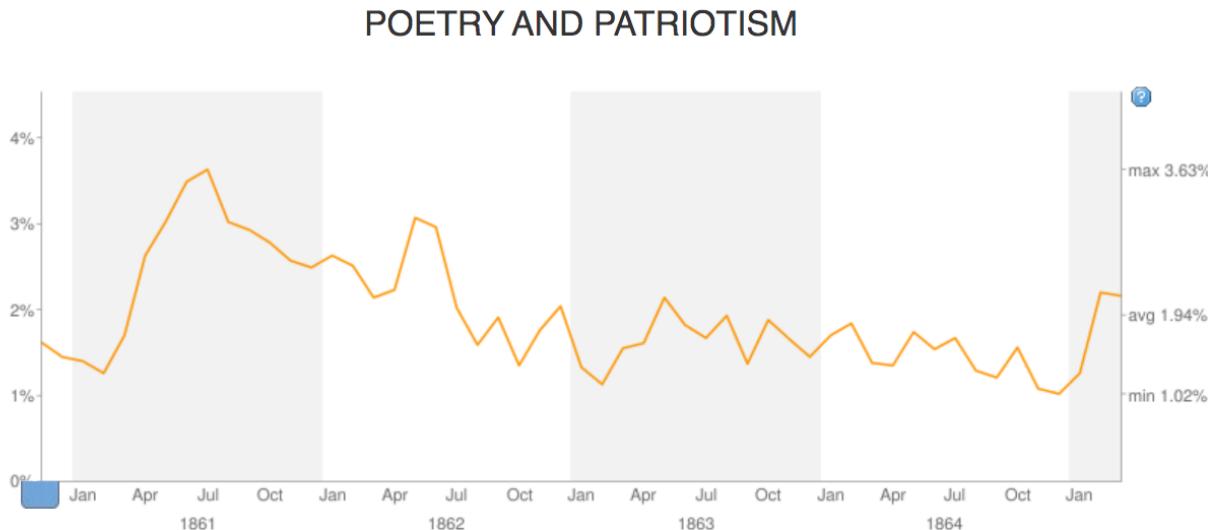


67%

«Обличения» отлично коррелируют с возвышенно-патриотической темой:



# Poetry and patriotism выглядит так:



## ADJUST CHART

threshold:  chart type:

## PREDICTIVE WORDS

COUNTRY LIFE GOD HEART BRAVE NOBLE  
HOME SOLDIER BATTLE MAN SPIRIT DEATH  
HONOR GALLANT SONS HOPE TRUE FIELD  
LOVE GLORIOUS HEARTS PATRIOTISM LAND  
DUTY

## EXEMPLARY ARTICLES

Thursday, March 30, 1865

"[Live as you will wish you had done when you come to die.](#)"

1



100%

Monday, November 25, 1861

[[Written for the Richmond Dispatch.](#)]

**GOD BLESS OUR BOYS!**

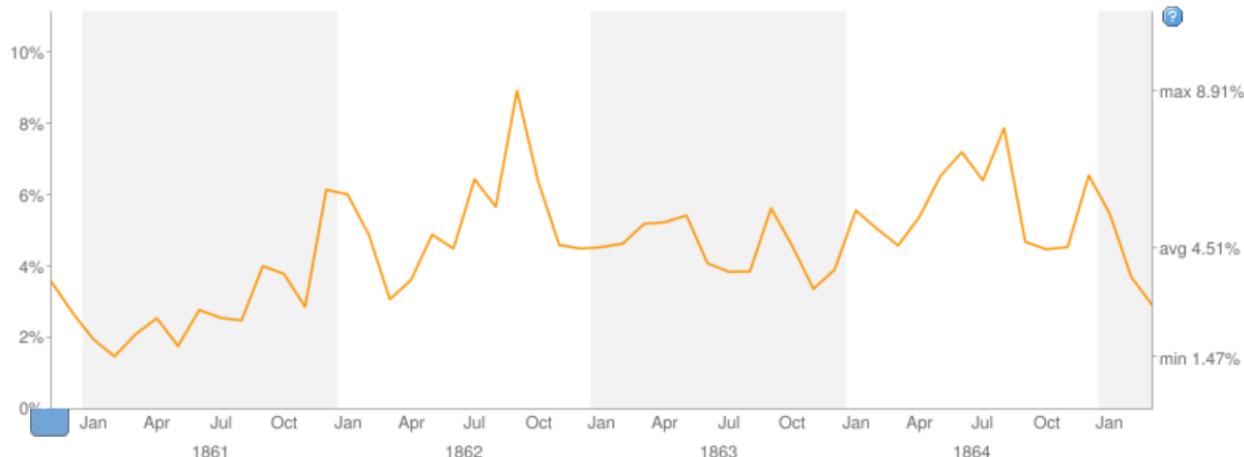
[By "L. A. M., a Richmond Girl.](#)



76%

# В отдельный топик выделились объявления о беглых рабах

FUGITIVE SLAVE ADS



ADJUST CHART

threshold:

chart type:

PREDICTIVE WORDS

NEGRO YEARS REWARD BOY MAN NAMED  
JAIL DELIVERY GIVE LEFT BLACK PAID PAY  
RAN COLOR RICHMOND SUBSCRIBER HIGH  
APPREHENSION AGE RANAWAY FREE FEET  
DELIVERED

EXEMPLARY ARTICLES

Monday, December 23, 1861

## RANAWAY--\$10 REWARD.

1

—Ranaway from the subscriber, on the 3d inst., my slave woman PARTHENA. Had on a dark brown and white calico dress. She is of a ginger-bread color; medium size; the right fore-finger shortened and crooked, from a whitlow. I think she is harbored somewhere in or . . .

[MORE](#)



90%

## RUNAWAYS.

**FIVE HUNDRED DOLLARS REWARD.**—Ran away from Greensboro', North Carolina, on the 15th of this month, a negro man, named LEWIS. Said negro is about twenty-four years old, five feet ten or eleven inches high, black complexion, and weighs one hundred and ninety-two pounds. He was bought on the 3d of February from Dr. R. H. Christian. I will pay the above reward for his delivery to me.

**ROBERT LUMPKIN,**  
Richmond, Virginia.

fe 16—1m\*

## Интерпретация графика:

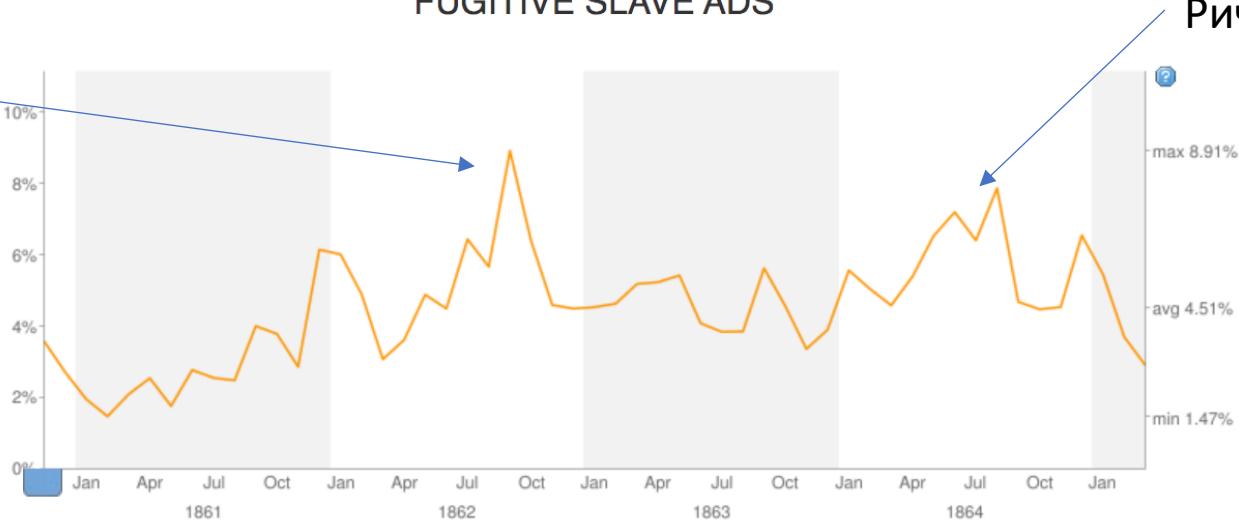
«For one, I think, this graph underscores the role of the Union army in presenting enslaved African Americans with opportunities—risky opportunities—to seize their freedom by running to the Yankee lines. There are two sustained spikes in the number of fugitive slave ads, the first in the summer of 1862 and the second in the summer of 1864. At both of those moments the Union army approached Richmond».

R. K. Nelson

# Два самых больших пика приходятся на приближение северян к городу:

Северяне первый раз приближаются к Ричмонду

FUGITIVE SLAVE ADS



ADJUST CHART

threshold: 0%

chart type: topic proportion/% print space

PREDICTIVE WORDS

NEGRO YEARS REWARD BOY MAN NAMED  
JAIL DELIVERY GIVE LEFT BLACK PAID PAY  
RAN COLOR RICHMOND SUBSCRIBER HIGH  
APPREHENSION AGE RANAWAY FREE FEET  
DELIVERED

EXEMPLARY ARTICLES

Monday, December 23, 1861

RANAWAY.--\$10 REWARD.

1

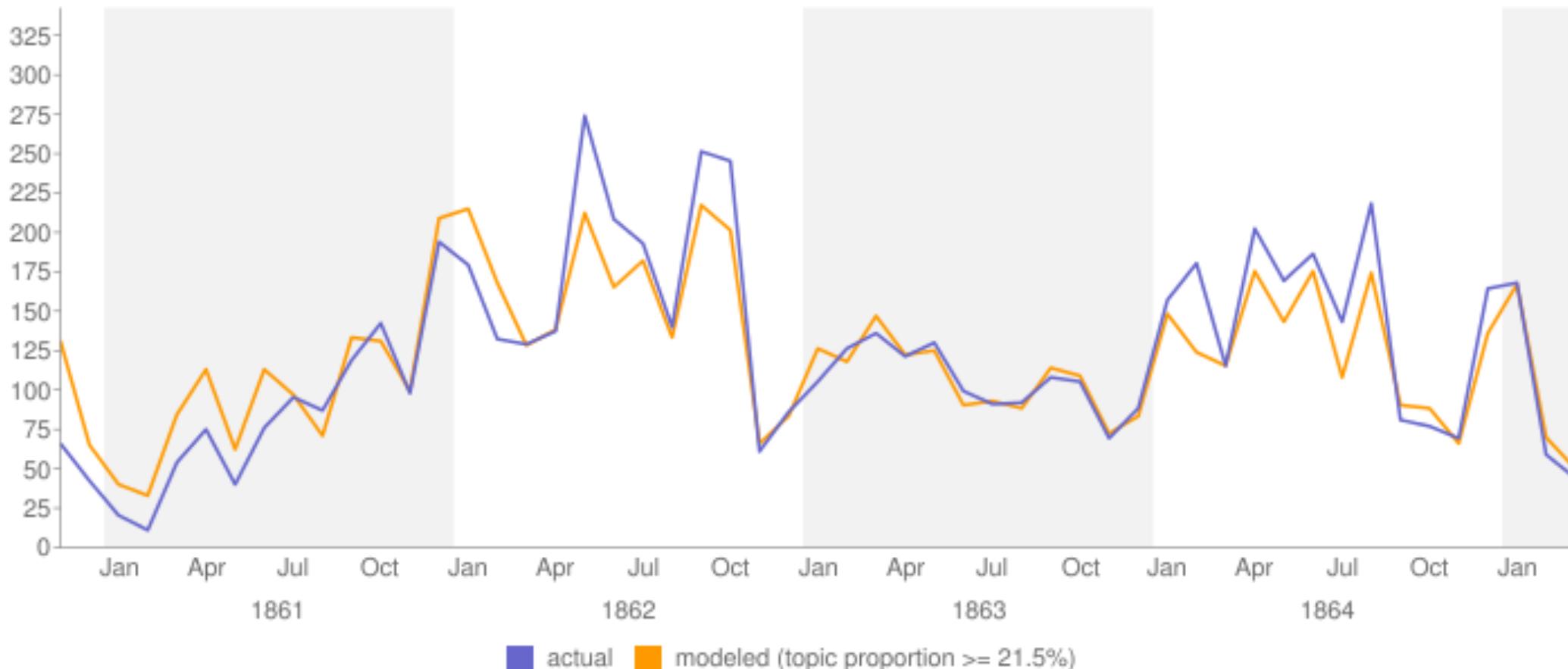
—Ranaway from the subscriber, on the 3d inst., my slave woman PARTHENA. Had on a dark brown and white calico dress. She is of a ginger-bread color; medium size; the right fore-finger shortened and crooked, from a whitlow. I think she is harbored somewhere in or . . .

MORE



90%

Выраженность топика коррелирует с реальным количеством объявлений:



# Циклы найма работников:

FOR HIRE AND WANTED ADS



## ADJUST CHART

threshold:

chart type:

## PREDICTIVE WORDS

STREET GOOD APPLY CORNER MAIN  
STREETS HOUSE JA CARY FRANKLIN BROAD  
STS ST HIRE COOK ROOM TS SERVANT FE  
WOMAN DE SERVANTS STORE YEAR

## EXEMPLARY ARTICLES

Wednesday, December 28, 1864

WANTED TO HIRE, a good COOK, WASHER and IRONER; by the month preferred; without encumbrance. Apply at No.

1



100%

193 Broad street.

de 28—3t\*

# Циклы найма работников:

## Начало войны, мобилизация



## ADJUST CHART

threshold: 0% 

chart type: topic proportion/% print space ▾

## PREDICTIVE WORDS

STREET GOOD APPLY CORNER MAIN  
STREETS HOUSE JA CARY FRANKLIN BROAD  
STS ST HIRE COOK ROOM TS SERVANT FE  
WOMAN DE SERVANTS STORE YEAR

## EXEMPLARY ARTICLES

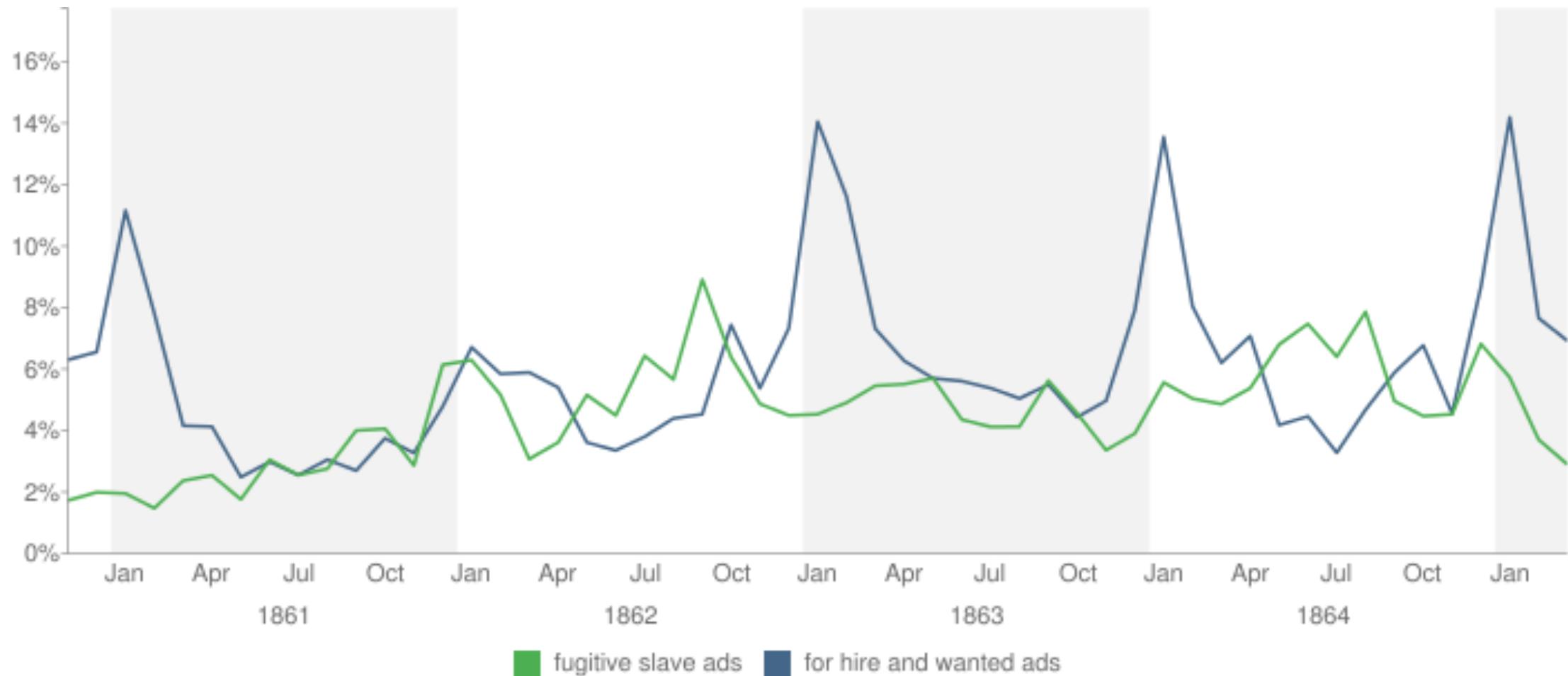
Wednesday, December 28, 1864

WANTED TO HIRE, a good COOK, WASHER and IRONER; by the month preferred; without encumbrance. Apply at No.

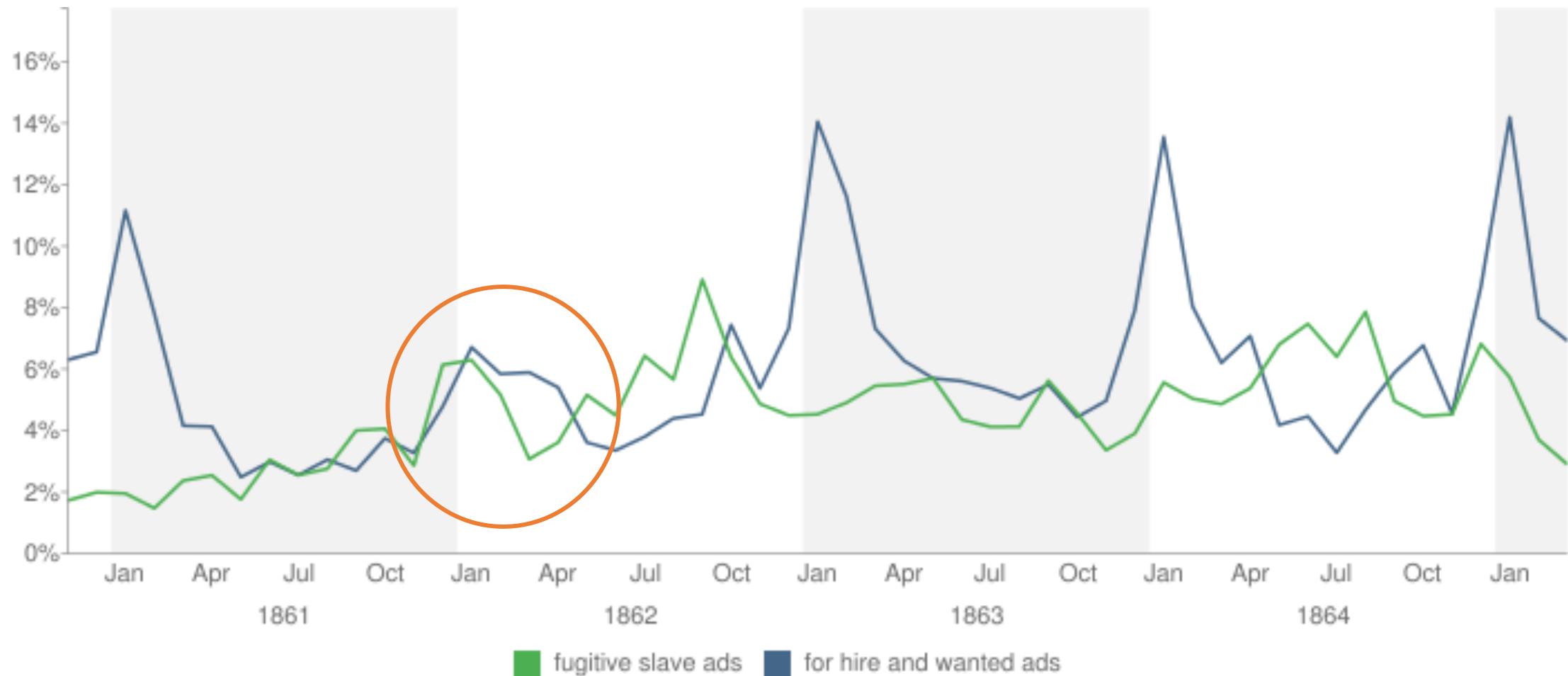
193 Broad street.

de 28—3t

В 1861 мобилизация все сломала — и, возможно, с этим связан и первый всплеск побегов тоже:



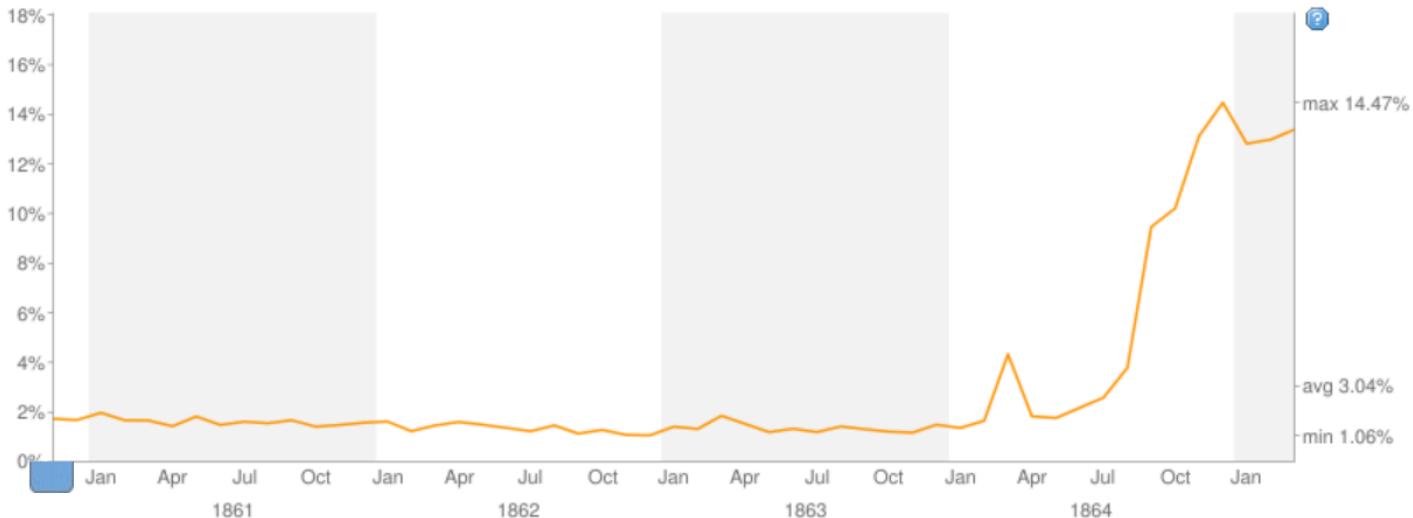
В 1861 мобилизация все сломала — и, возможно, с этим связан и первый всплеск побегов рабов:



А еще

# Экономика

## WAR BONDS



### ADJUST CHART

threshold:

chart type:

### PREDICTIVE WORDS

TREASURY BONDS NOTES SECRETARY  
INTEREST CENT ACT JANUARY BANK  
EXCHANGE ST PAYMENT CERTIFICATES DAY  
ISSUED PAY BANKS TREASURER AMOUNT  
RECEIVE ISSUE AUTHORIZED DEPOSITARIES  
CONFEDERATE

### EXEMPLARY ARTICLES

Friday, September 02, 1864

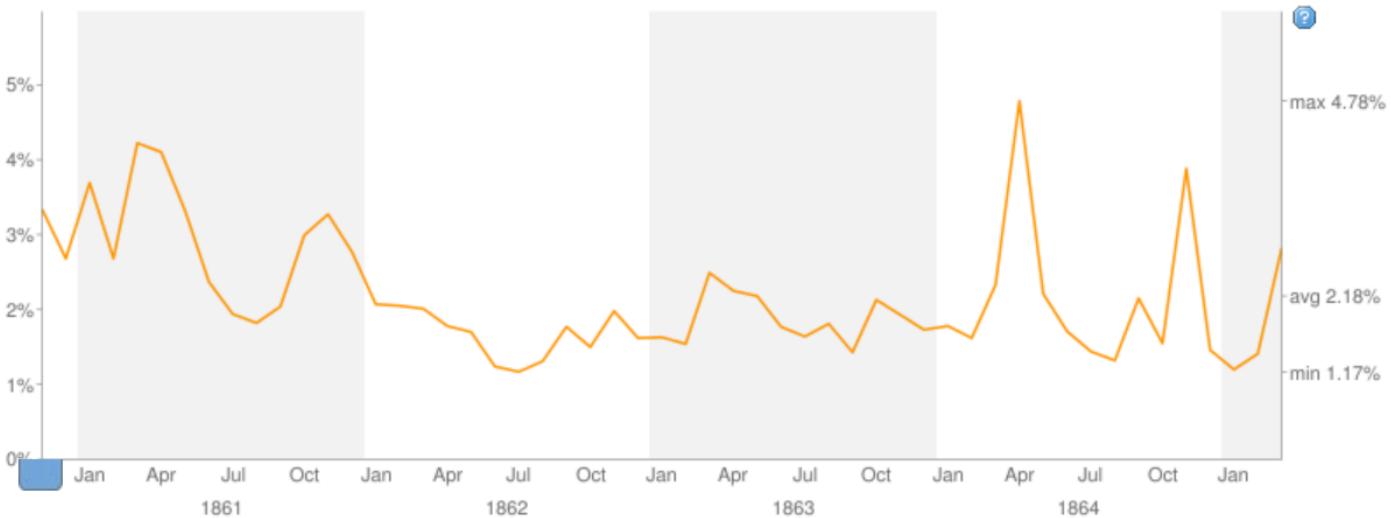
1 [REGULATIONS FOR CARRYING INTO EFFECT THE ACT OF CONGRESS APPROVED JUNE 13, 1864, AUTHORIZING OWNERS OF REGISTERED EIGHT PER CENT BONDS, ISSUED UNDER THE PROVISIONS OF THE ACT APPROVED MAY 16, 1861, TO EXCHANGE THE SAME FOR COUPON BONDS.—Parties owning registered eight per cent. bonds under the act of . . .](#) [MORE](#)



100%

# Политика

## ELECTIONS



### ADJUST CHART

threshold:

chart type:

### PREDICTIVE WORDS

VIRGINIA STATE GOVERNOR DAY  
CONVENTION CITY ELECTION  
COMMONWEALTH HELD MEETING VOTE  
ELECTED DISTRICT MEMBERS COUNCIL  
LEGISLATURE APPOINTED MAJORITY  
ORDINANCE JOHN COUNTIES HALL

### EXEMPLARY ARTICLES

Monday, August 10, 1863

NOTICE TO TAX-PAYERS.

1 —State of Virginia, 12th Collector's ...



100%

Thursday, November 21, 1861

BY THE GOVERNOR OF VIRGINIA.--A PROCLAMATION.

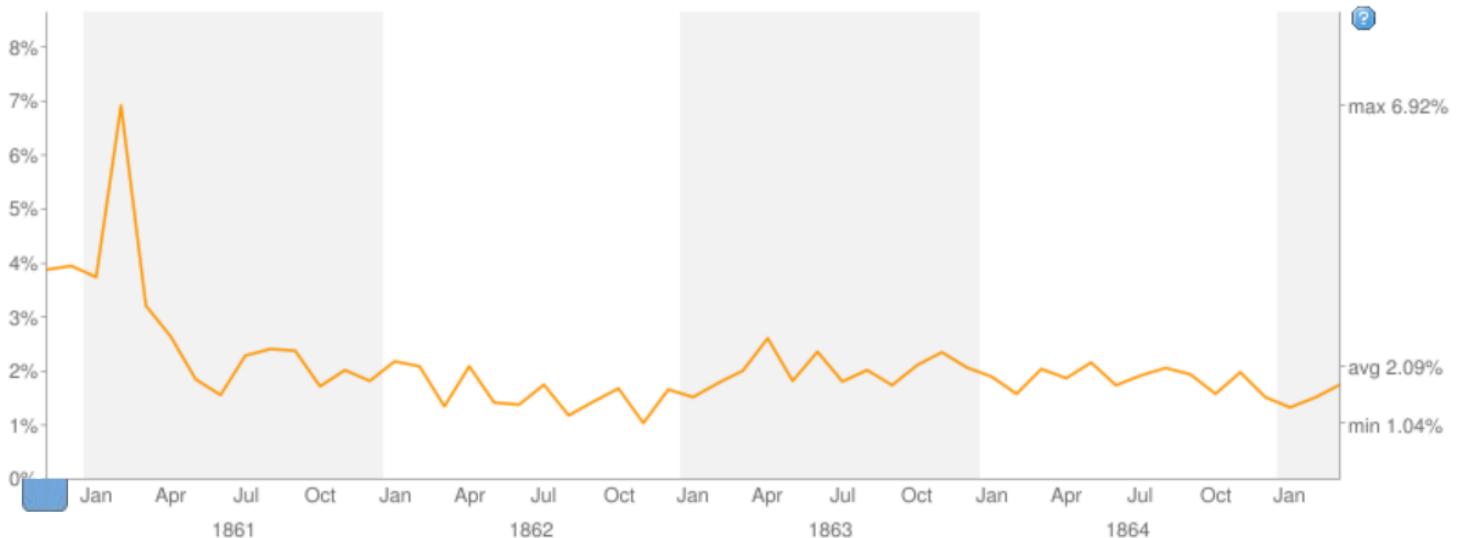


94%

—Whereas, a vacancy has occurred, in the representation of this

# Повседневность

## GROCERIES



### ADJUST CHART

threshold:

chart type:

### EXEMPLARY ARTICLES ?

Monday, May 13, 1861

No. 1 LARD.

### PREDICTIVE WORDS ?

BD BANK CENTS MARKET SALES TOBACCO  
SALE STOCK FLOUR CORN BBLS SUGAR CTS  
VIRGINIA COFFEE WHEAT OIL FIRM WHITE  
BACON DULL GOOD BC BUSHEL

1 —10 bbls., for sale by

my 8 I. & G. B. DAVENPORT.



100%

# Спасибо за внимание

Окончание следует