

Case study: Comcast Consumer reviews

Olabanji Shonibare and Melissa Davidson

March 7, 2017

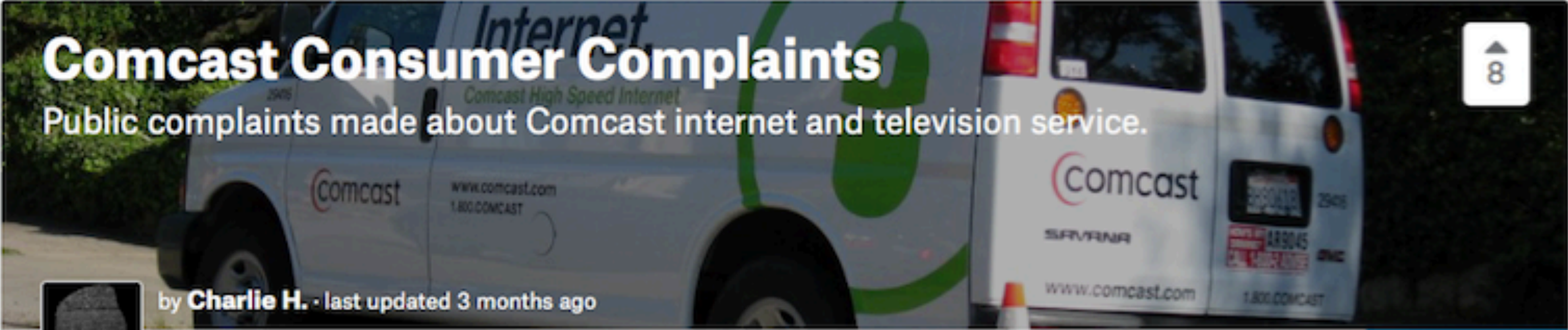
Outline

- Introduction
- Word co-occurrences
- Topic modeling
- Sentiment analysis

Introduction

Kaggle

kaggle Q [Competitions](#) [Datasets](#) [Kernels](#) [Discussion](#) [Jobs](#) [Sign Up](#) [Log In](#)



Comcast Consumer Complaints

Public complaints made about Comcast internet and television service.

by **Charlie H.** · last updated 3 months ago

[Overview](#) [Kernels](#) [Discussion](#) [Activity](#) [Download \(4 MB\)](#) [New Notebook](#) [New Script](#)




Kernels

first run run 3 months ago	4 votes
Top 10 Customer Locations/... run 3 months ago	3 votes
Notebook949b289ba3 run 22 days ago	0 votes

Discussion

There are no conversations yet.
[Start one](#)

Top Contributors

 Dan	1st
 ne0j0y	2nd
 Ola	3rd

Comcast Consumer Complaints

- comcast_consumeraffairs_complaints.csv

Raw complaint data about Comcast television and internet published at **consumeraffairs.com** between **04/08** and **09/16**.

- comcast_fcc_complaints_2015.csv

Raw complaints made to the **FCC** about Comcast between **04/15** and **06/15**.

Consumer Affairs



author	posted_on	rating	text			
Alantae of Ch	Nov. 22, 201	1	I used to love Comcast. Until all these constant upd			
Vera of Phila	Nov. 19, 201	1	I'm so over Comcast! The worst internet provider. I			
Sarah of Ran	Nov. 17, 201	1	If I could give them a negative star or no stars on th			
Dennis of Ma	Nov. 16, 201	1	I've had the worst experiences so far since install or			


author	posted_on	rating	text			
16-Nov	2008 technik	wires and ga	2008 between 2-5 p.m.			
16-Nov	2008 technik	wires and ga	2008 between 2-5 p.m.			
16-Nov	2008 technik	wires and ga	2008 between 2-5 p.m.			
Tech line call	but Charles	the word dir	but he is not	2008At som	an amplifier	weak signals
Tech line call	but Charles	the word dir	but he is not	2008Around	2008	technician d
Tech line call	but Charles	the word dir	but he is not	2008Novem	2008 Integra	removed am
Tech line call	but Charles	the word dir	but he is not	2008Novem	2008 technik	wires and ga

Consumer Affairs

CONSUMERAFFAIRS

Consumer NewsBuyers GuidesFor Businesses

  Write a review

 Search

Log in

STEP 1STEP 2STEP 3

Let's get started

Login

Create account

Why should I create an account?

About ConsumerAffairs

ConsumerAffairs is a private, non-governmental entity that empowers consumers by providing a forum for their complaints and a means to be contacted by lawyers if their complaints have legal merit. Your complaints and reviews may be published, shared with the news media and reviewed by attorneys at no cost to you. See our [FAQ](#) for more information

FCC

Ticket #	Customer Co	Date	Time	Received Via	City	State	Zip code	Status	Filing on Beh	Description		
250635	Comcast Cab	4/22/15	3:53:50 PM	Internet	Abingdon	Maryland	21009	Closed	No	I have been contacting Comcast Internet Tech		
223441	Payment disa	4/8/15	#####	Internet	Acworth	Georgia	30102	Closed	No	Back in January 2015 I made 2 payments: One		
242732	Speed and Se	4/18/15	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes	Our home is located at in Acworth Georgia 30		
277946	Comcast Imp	5/7/15	#####	Internet	Acworth	Georgia	30101	Open	Yes	Comcast in the Atlanta area has just put into e		
307175	Comcast not	5/26/15	1:25:26 PM	Internet	Acworth	Georgia	30101	Solved	No	I have been a customer of Comcast of some sc		

[Sign in](#)[Check Status](#)

Consumer Complaint Center

Internet Complaint

Approved by OMB 3060-0874 (Est. average burden per response is 15 minutes).

[Privacy Statement](#)

Fields marked with * are required.

The submit button will not be active until all required fields are completed.

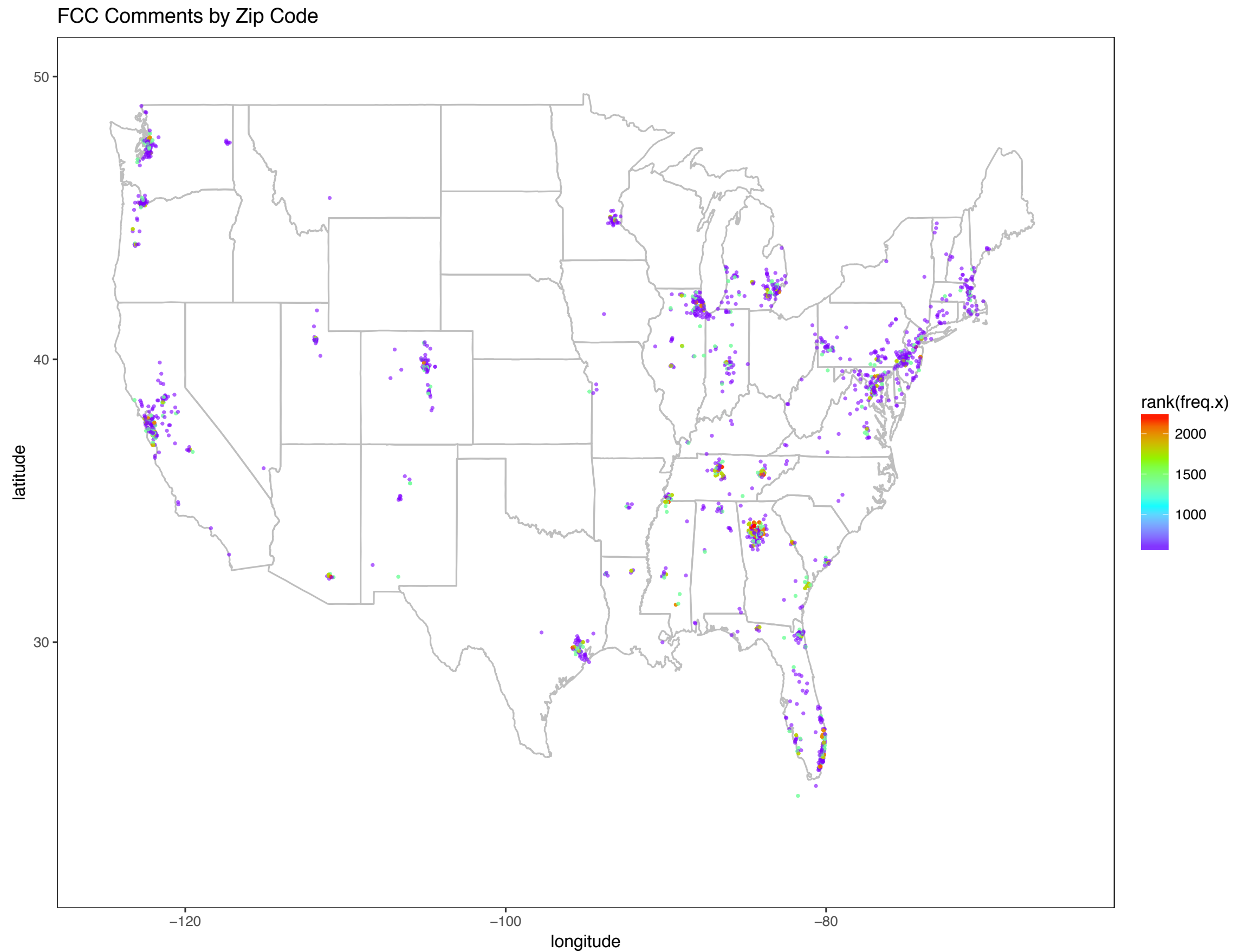
Web browsers supported: Google Chrome, Mozilla Firefox, and Apple Safari (latest two versions) and Microsoft Internet Explorer (latest version only).

Your email address *

Subject *

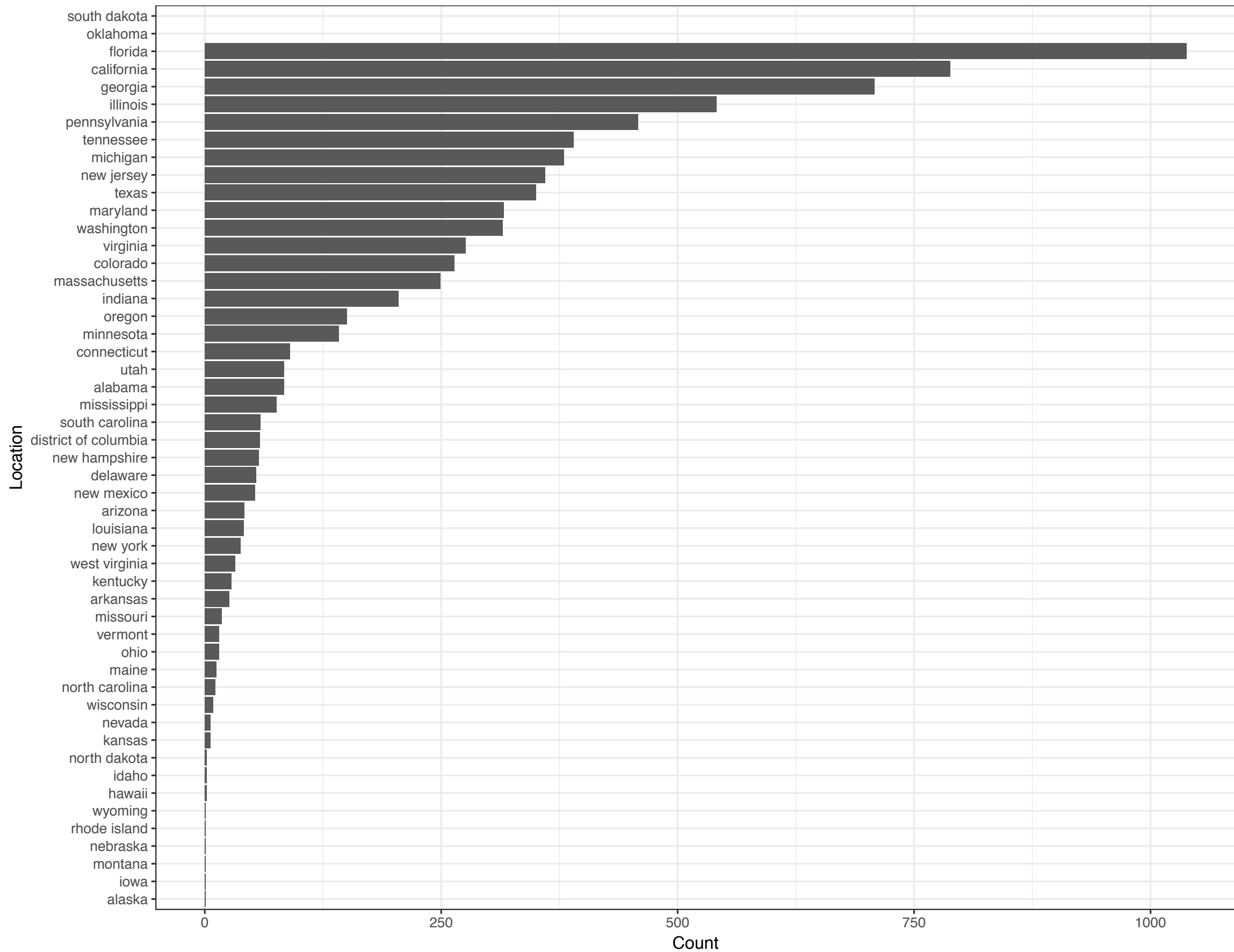


Complaints by zip code (FCC)



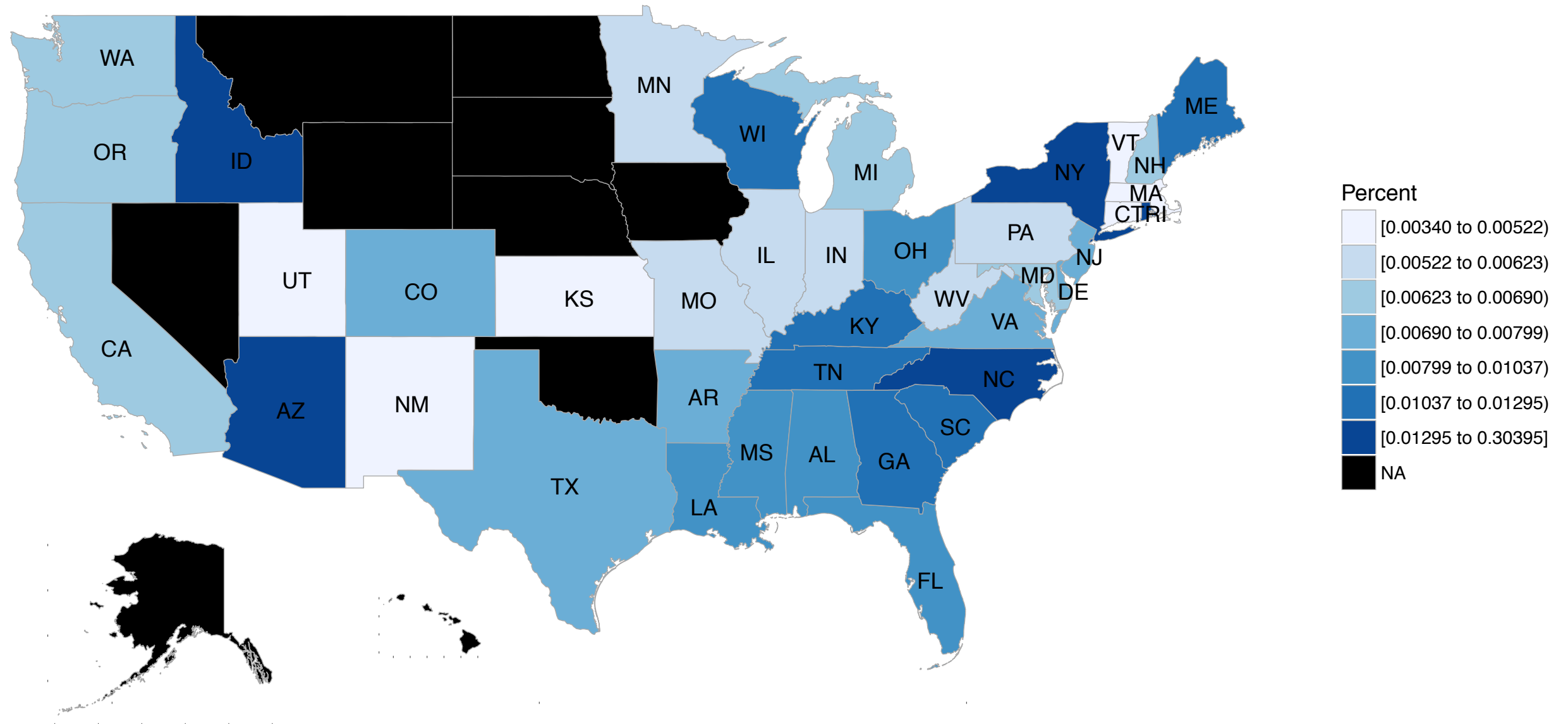
CA and FCC

Complaints



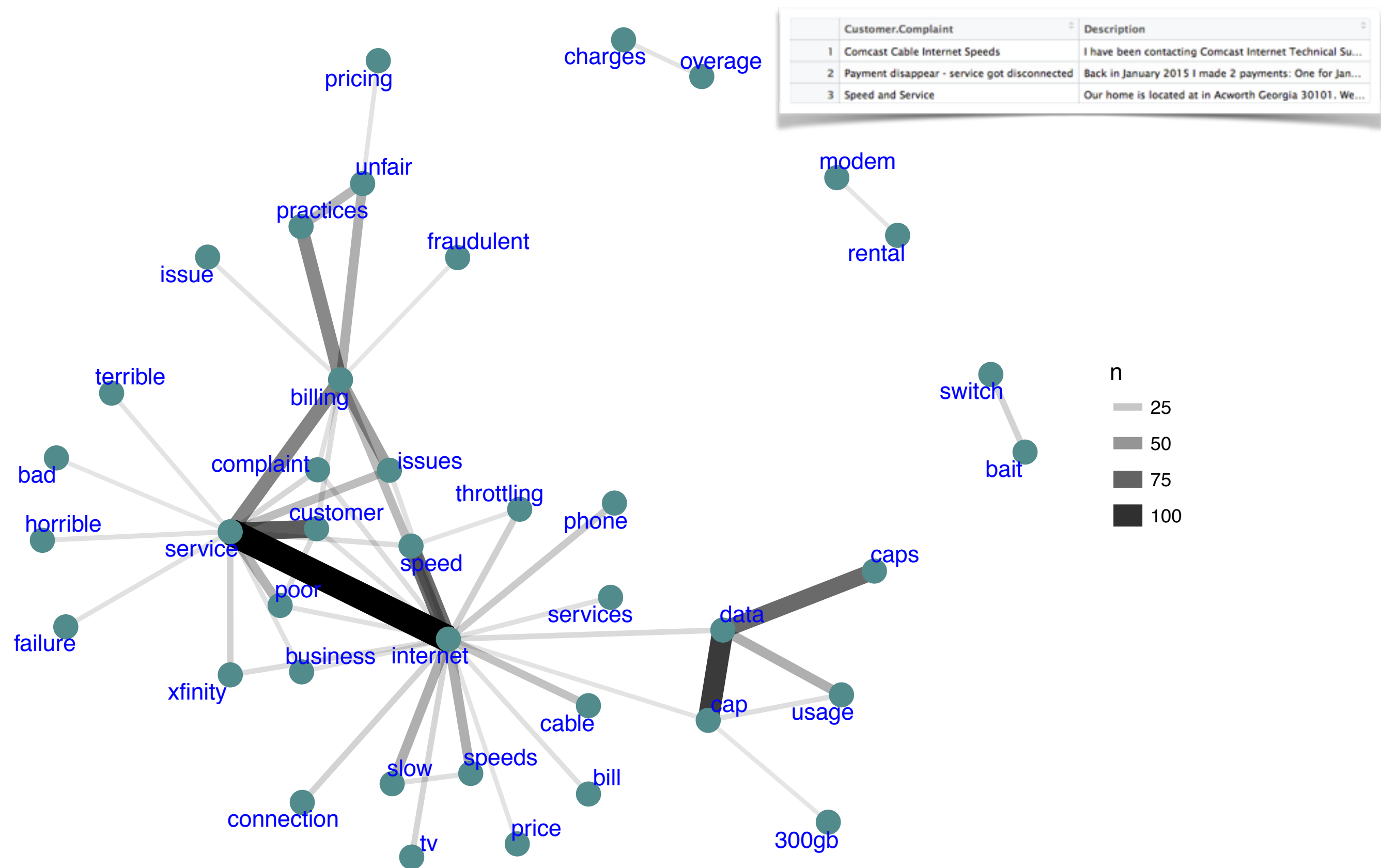
Complaints by percentage

Both Percentage for Areas with Comcast

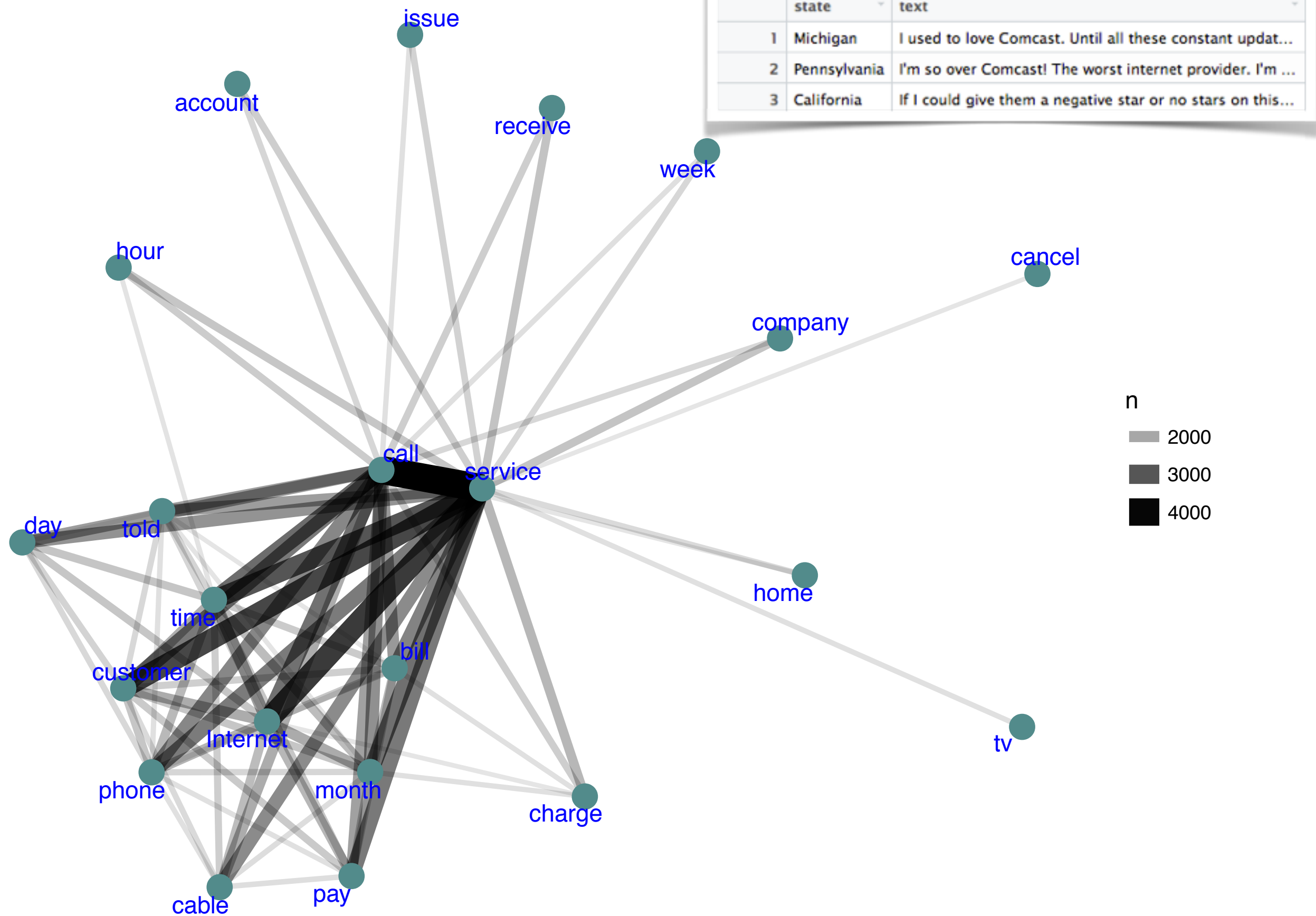


Word co-occurrences

Co-occurrence network in Comcast dataset titles (FCC)



Co-occurrence network in Comcast dataset text



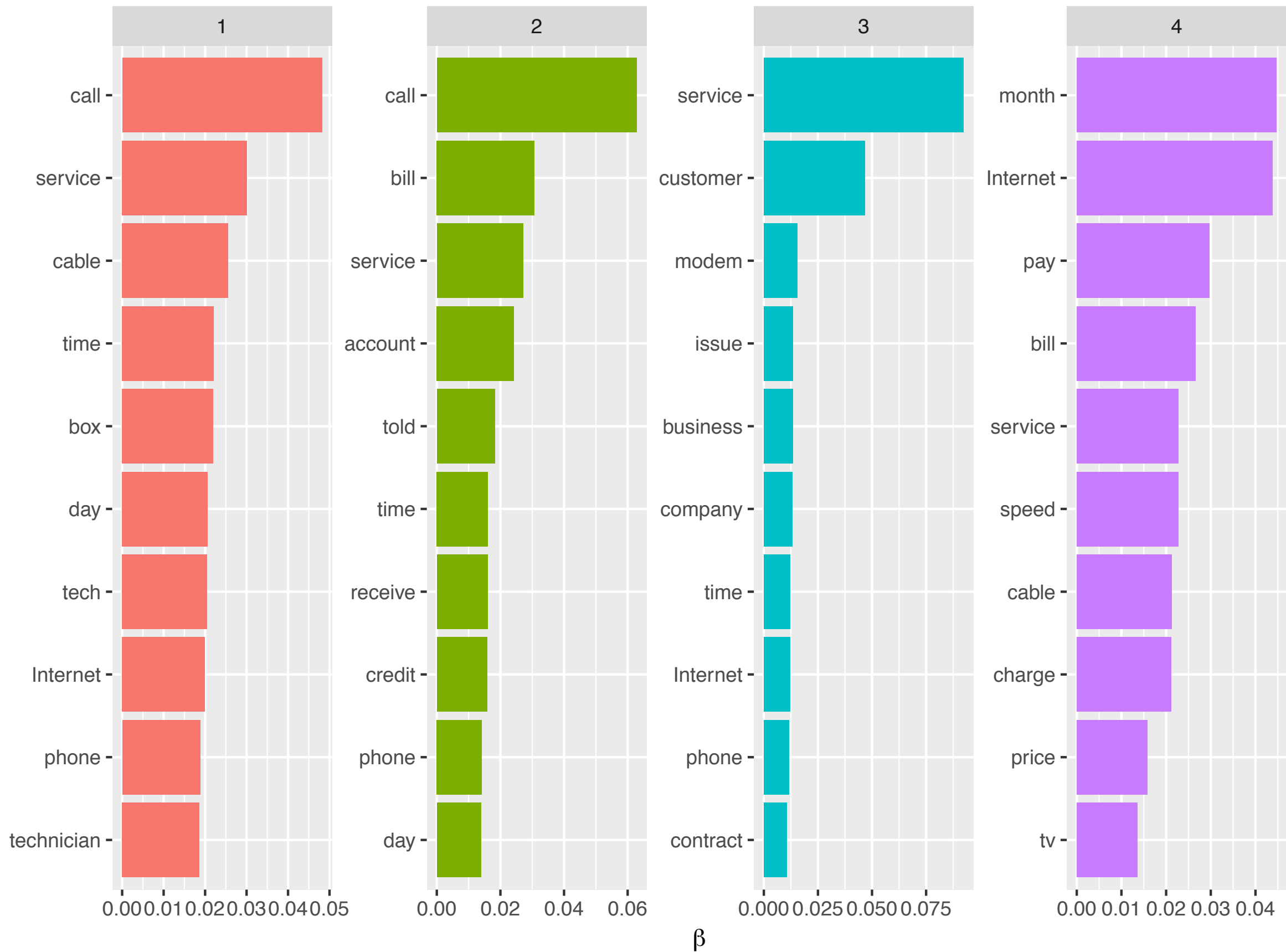
Topic modeling

Topic modeling

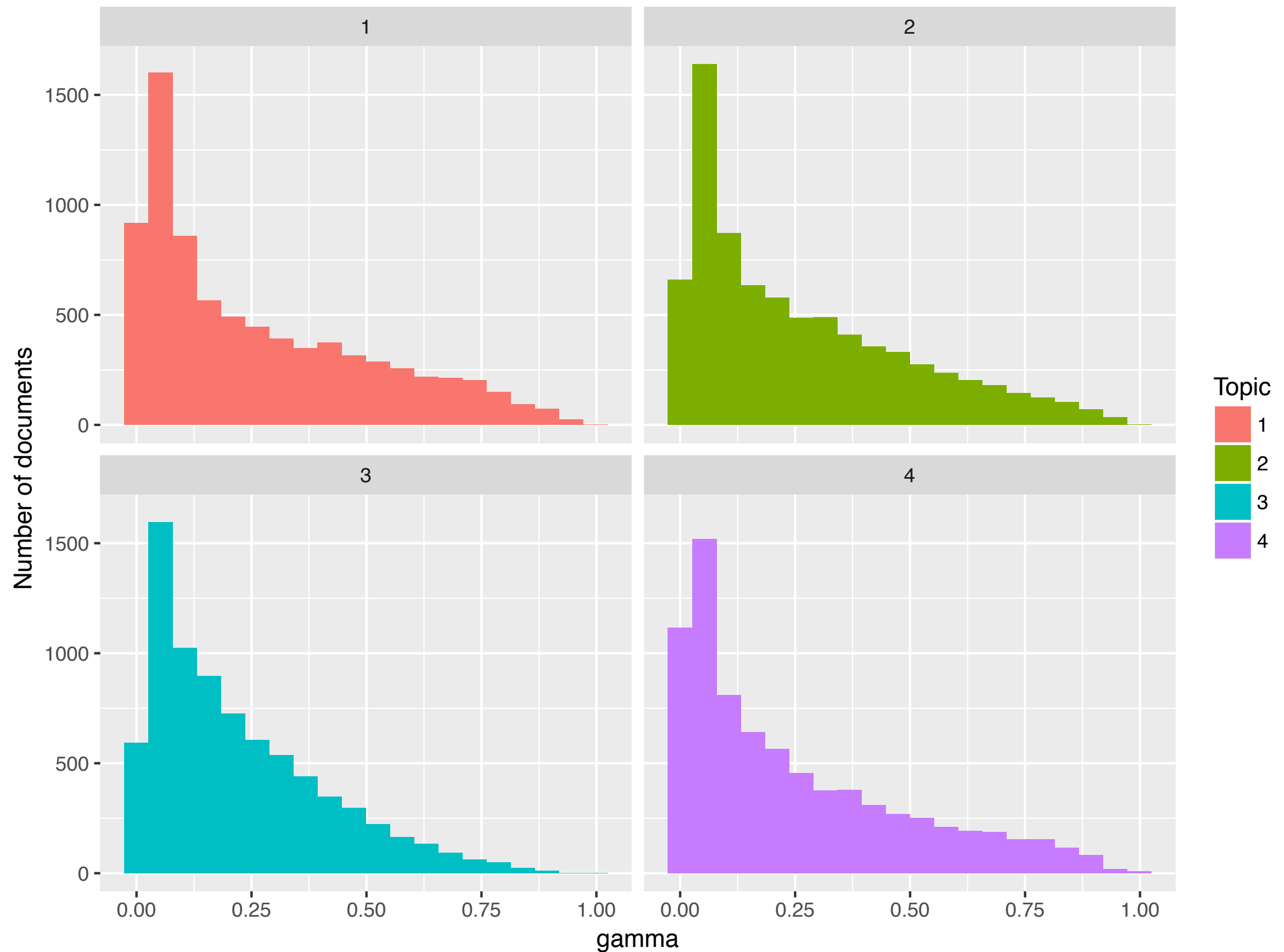
Topic modeling is a method for unsupervised classification of documents, similar to “clustering” on numerical data.

Topic model: Latent Dirichlet Allocation (LDA)

Top 10 terms in each LDA topic



Probability distribution for each topic (**FCC**)



Sentiment analysis

Sentiment analysis

It is a computational approach to identify people's opinion towards an entity

Classification techniques

- Machine learning
- Lexicon-based

Sentiment analysis

Classification techniques

- Machine learning
- Lexicon-based

NRC

```
## # A tibble: 13,901 × 2
##       word sentiment
##       <chr>      <chr>
## 1    abacus      trust
## 2   abandon     fear
## 3   abandon  negative
## 4   abandon  sadness
## 5  abandoned    anger
## 6  abandoned    fear
## 7  abandoned  negative
## 8  abandoned  sadness
## 9 abandonment  anger
## 10 abandonment  fear
## # ... with 13,891 more rows
```

BING

```
## # A tibble: 6,788 × 2
##       word sentiment
##       <chr>      <chr>
## 1    2-faced  negative
## 2    2-faces  negative
## 3         a+   positive
## 4   abnormal  negative
## 5   abolish  negative
## 6  abominable  negative
## 7  abominably  negative
## 8  abominate  negative
## 9  abomination  negative
## 10      abort  negative
## # ... with 6,778 more rows
```

AFINN

```
## # A tibble: 2,476 × 2
##       word score
##       <chr> <int>
## 1   abandon    -2
## 2  abandoned    -2
## 3  abandons    -2
## 4  abducted    -2
## 5  abduction    -2
## 6  abductions    -2
## 7    abhor     -3
## 8  abhorred    -3
## 9  abhorrent    -3
## 10   abhors     -3
## # ... with 2,466 more rows
```

AFINN lexicon

- This is utterly *excellent*!

Sentiment score = 3

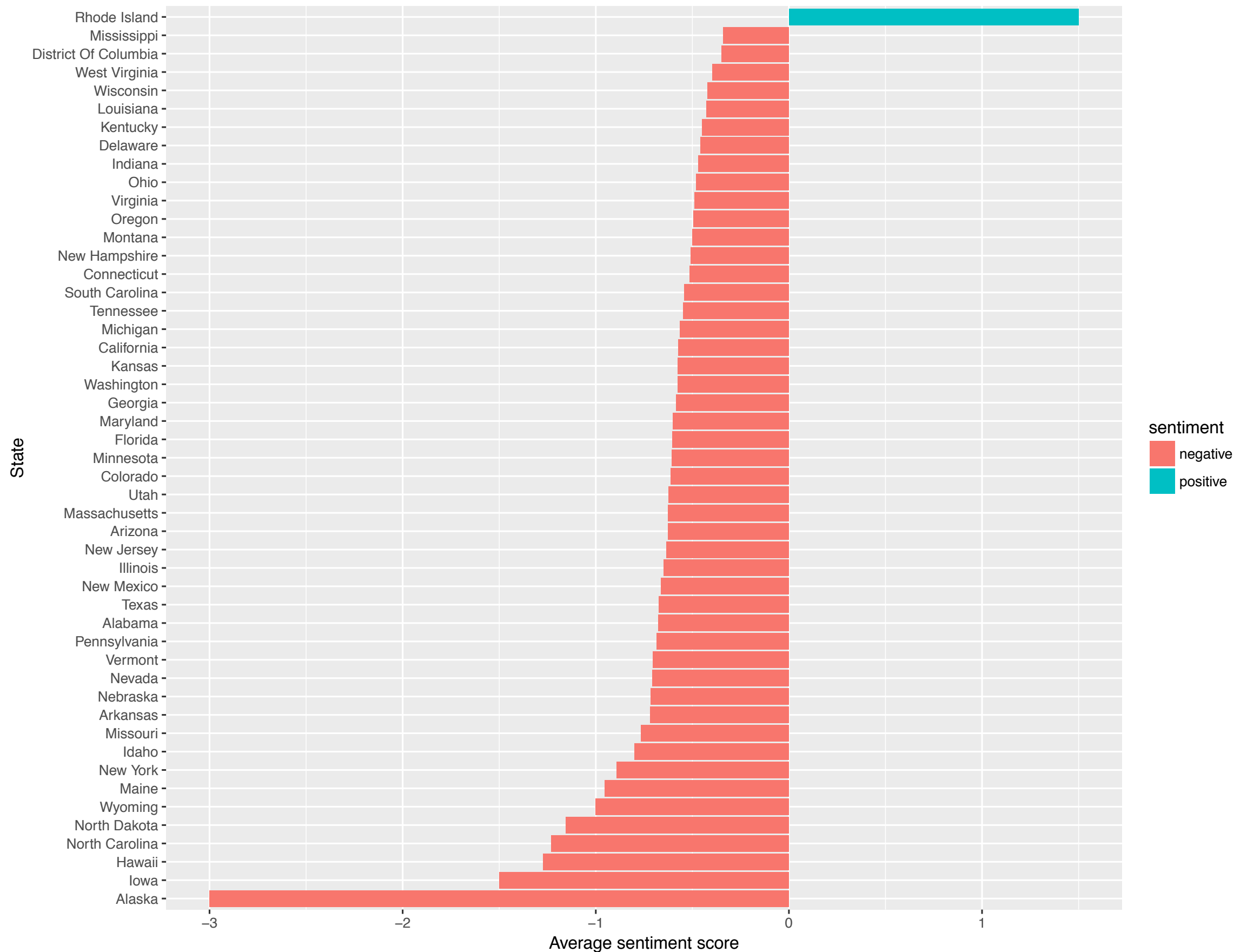
- "I continue to receive *unwanted* calls from Comcast despite my instructions."

Sentiment score = -2

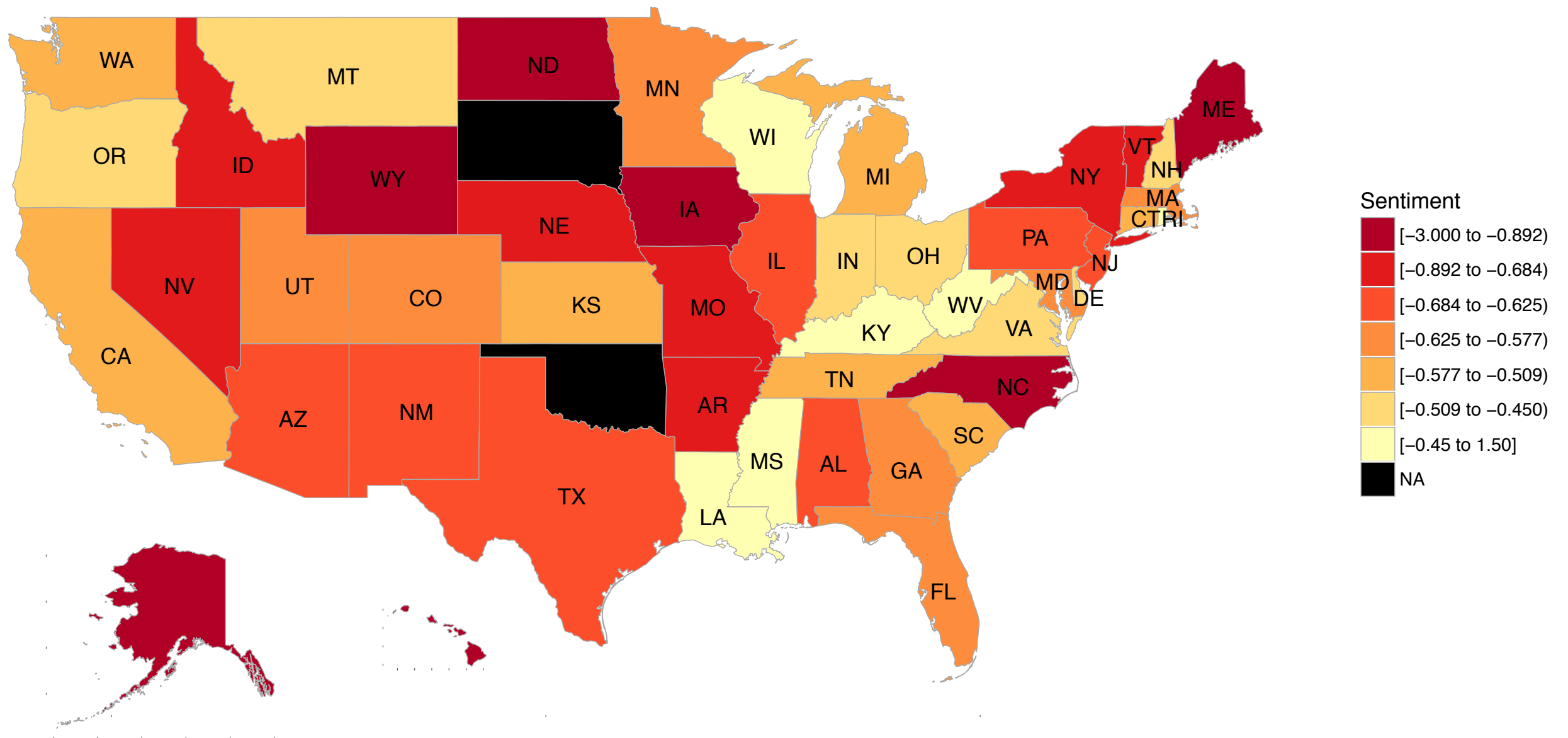
- " I'm *not happy* and I *don't like* it "

Sentiment score = 5

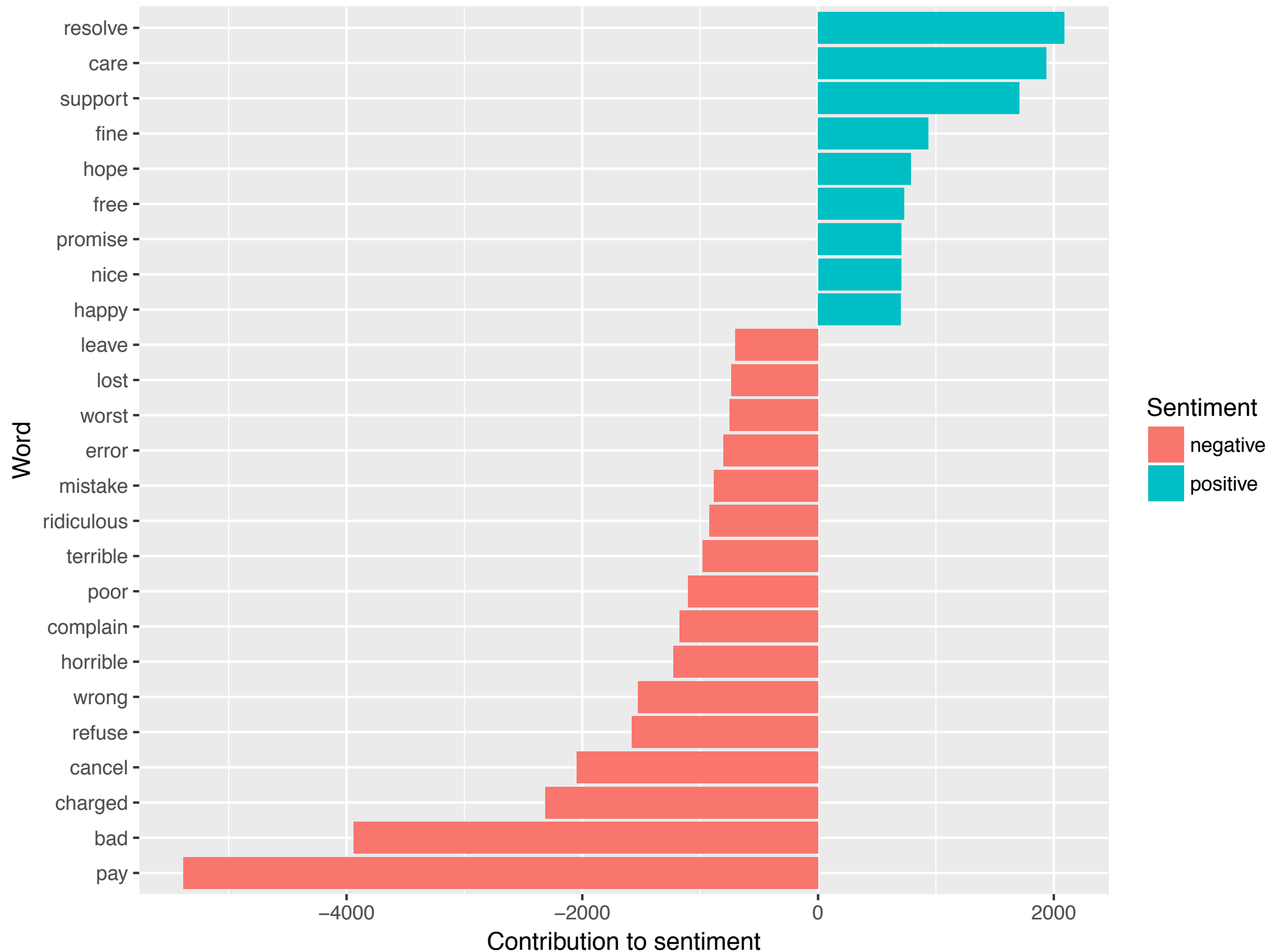
Average AFINN score for reviews within each state



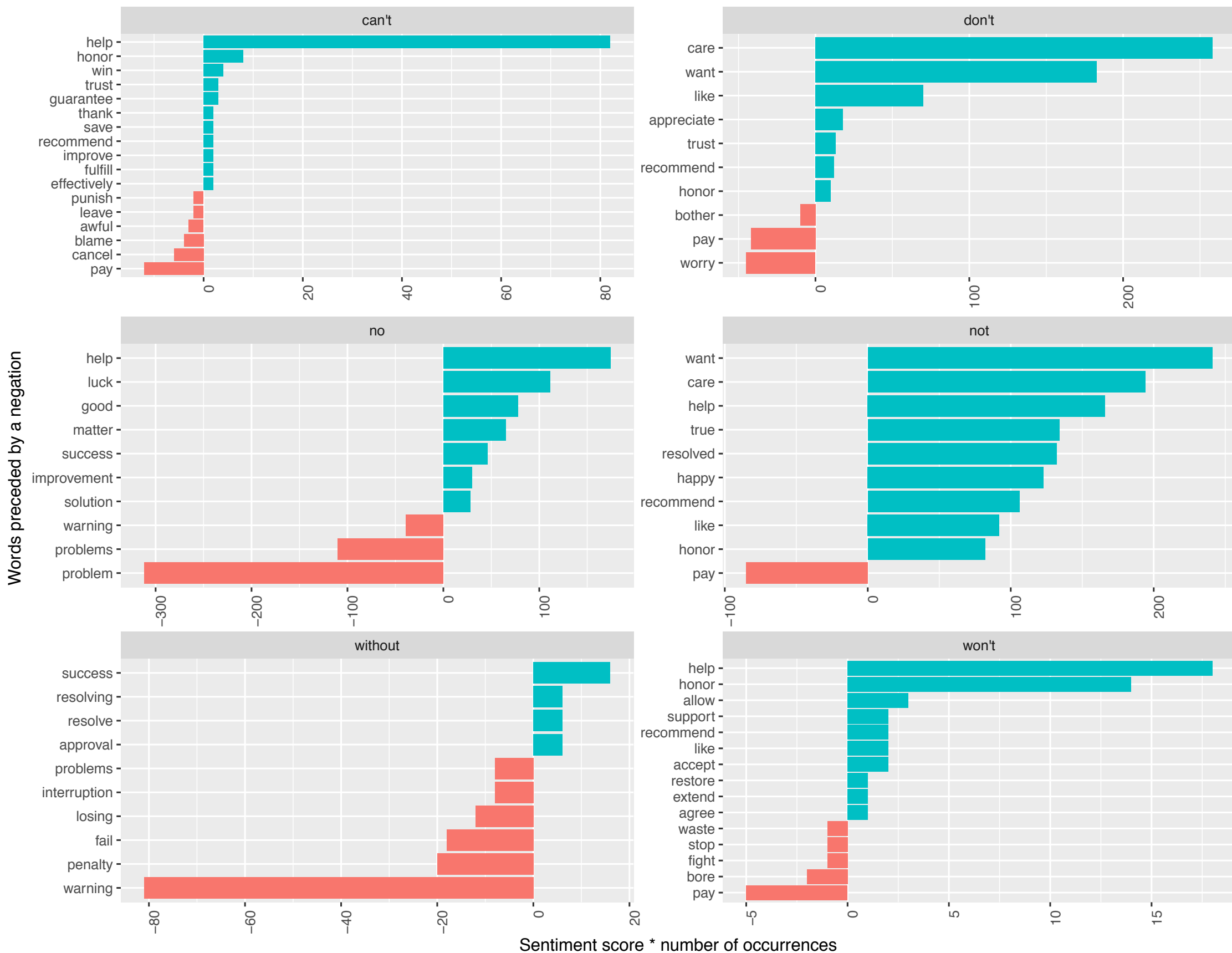
Average AFINN score for reviews within each state



Words with the highest contribution to sentiment scores



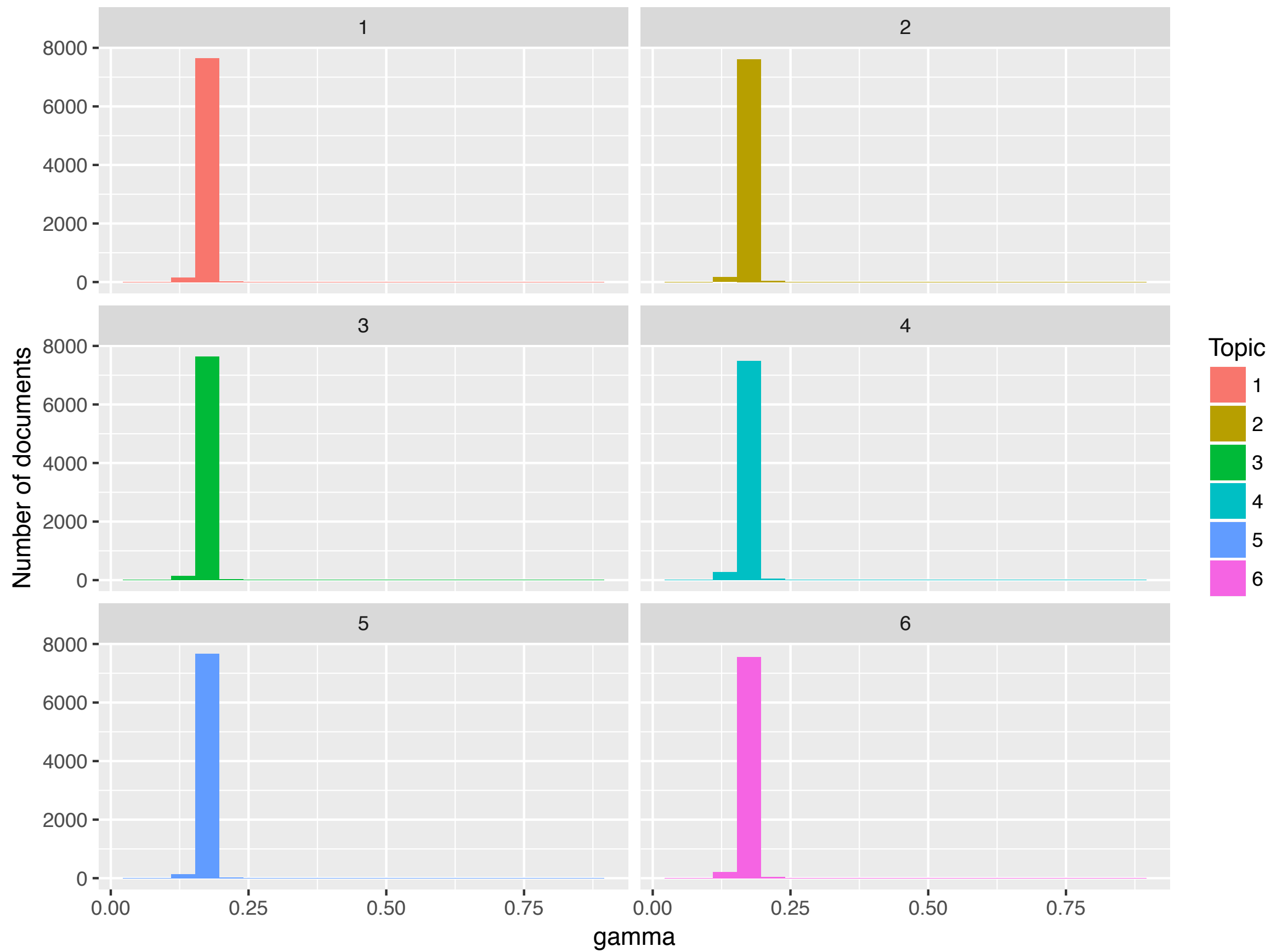
Negations that contributed the most to sentiment (**FCC**)



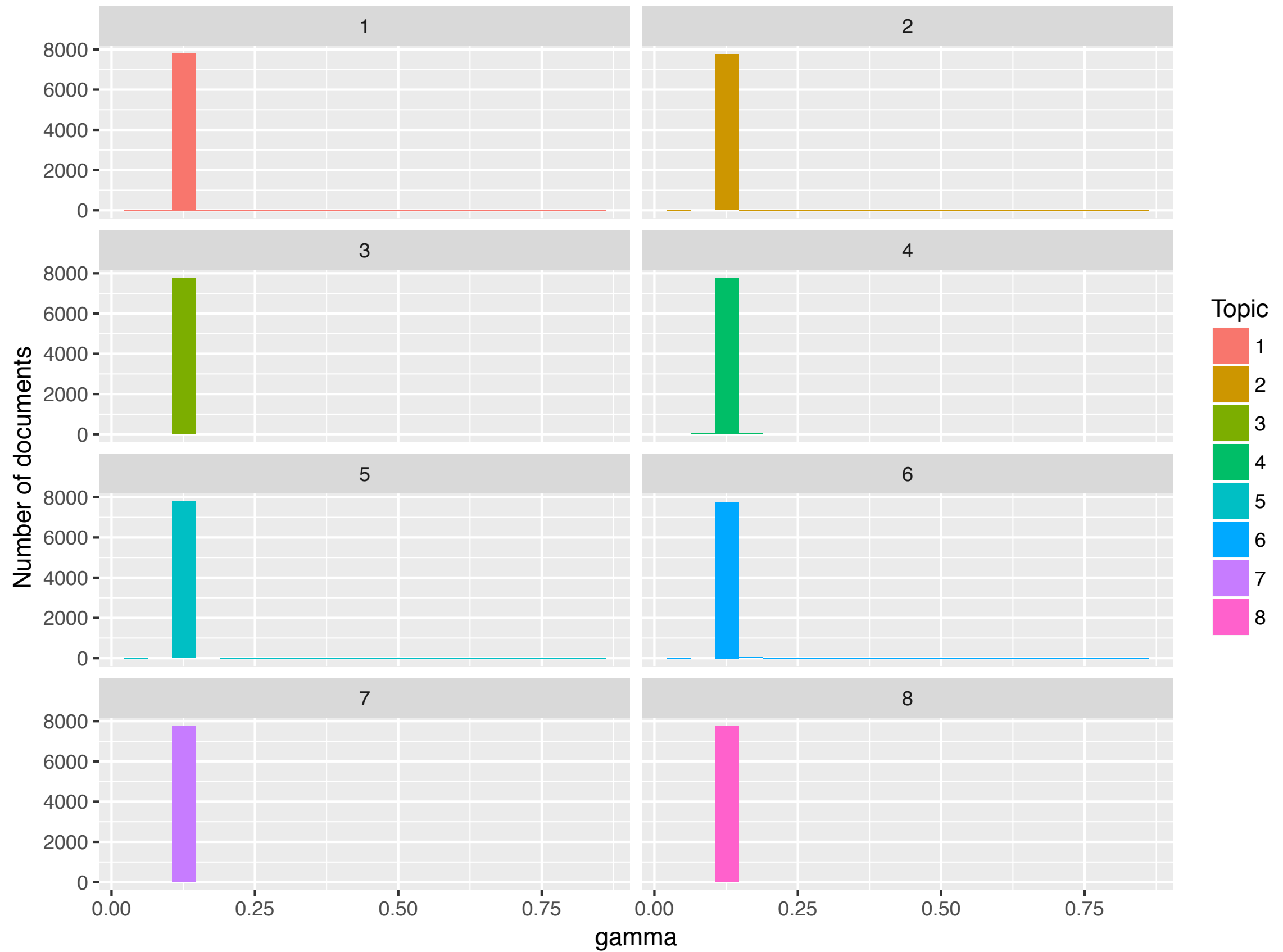
Thank you

Questions or comments?

Number of topics = 6



Number of topics = 8



Number of topics = 12

