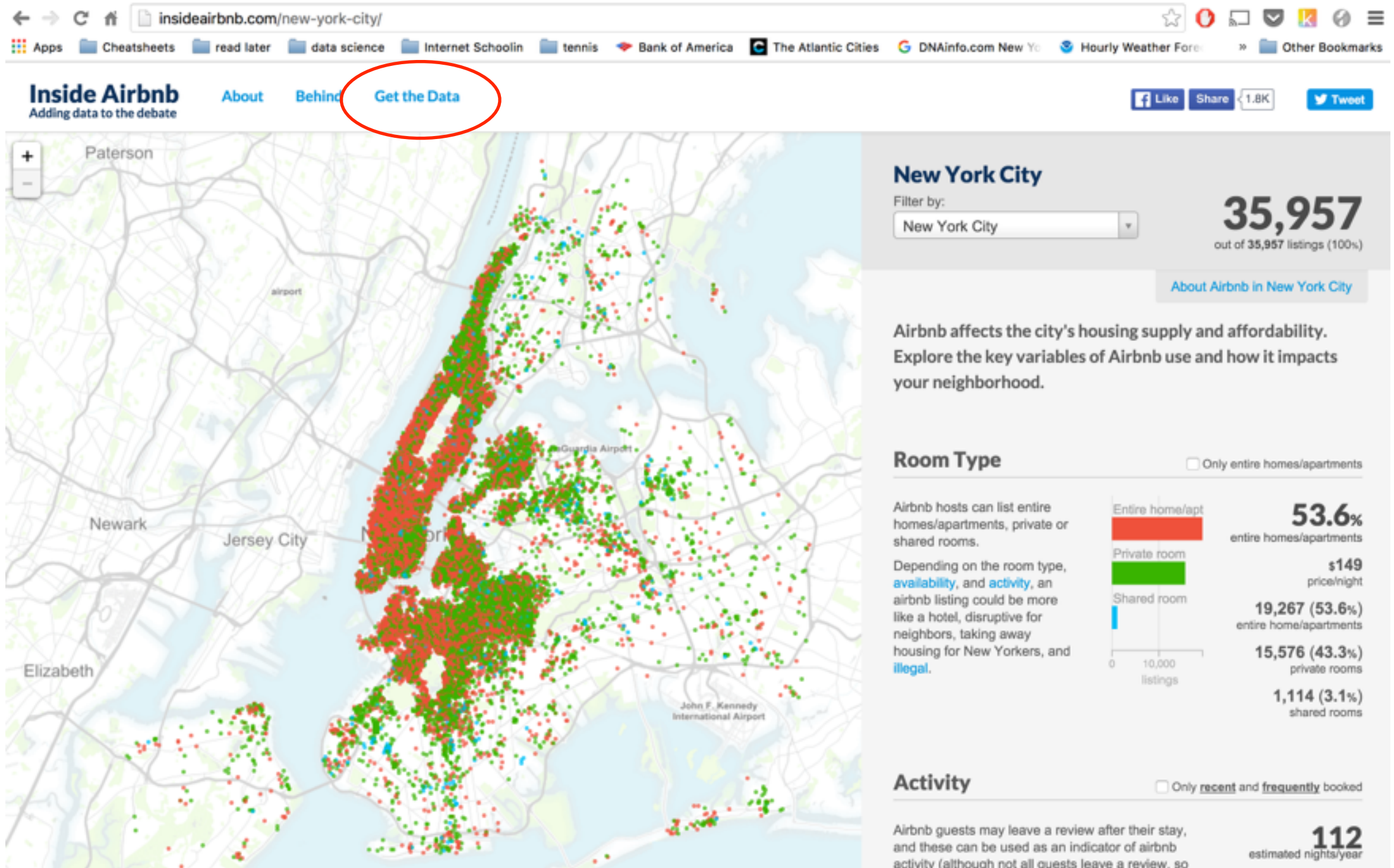


# Most Commonly Used Words in Airbnb Reviews, by Neighborhood

Dan Yawitz  
[github.com/yawitzd](https://github.com/yawitzd)  
March 4, 2016

# Getting the data



# Getting the data

35,597 airbnb listings

439,092 airbnb reviews, 2008-2016

# Getting the data

35,597 airbnb listings

439,092 airbnb reviews, 2008-2016

```
In [17]: auto = 'The host canceled this reservation | days before arrival | This is an automated posting'
r = r[r.comments.str.contains(auto) == False]
```

# Getting the data

35,597 airbnb listings

420,065 airbnb reviews, 2008-2016

```
In [17]: auto = 'The host canceled this reservation | days before arrival | This is an automated posting'
r = r[r.comments.str.contains(auto) == False]
```

# Removing stop words

```
In [15]: foreign = ['est', 'et', 'es', 'de', 'que', 'la', 'le', 'much', 'bien', 'pour', 'un', 'en', \
                    'appartement', 're', 'un', 'und', 'nous', 'tres', 'el', 'pas', 'muy', 'vie', 'las', 'los', 'die',
                    'ich', 'ein', 'die', 'nos', 'ich', 'su', 'ist', 'wir', 'para', 'il', 'chi']
```

```
In [16]: nyc = ['nyc', 'brooklyn', 'thanks', 'great', 'located', 'train', 'subway', 'new york', \
                'ny', 'york', 'nights', 'mrs', 'ms',
                ]
```

```
In [18]: full_stop = stop + names + nyc + foreign
```

# TF-IDF Weighting

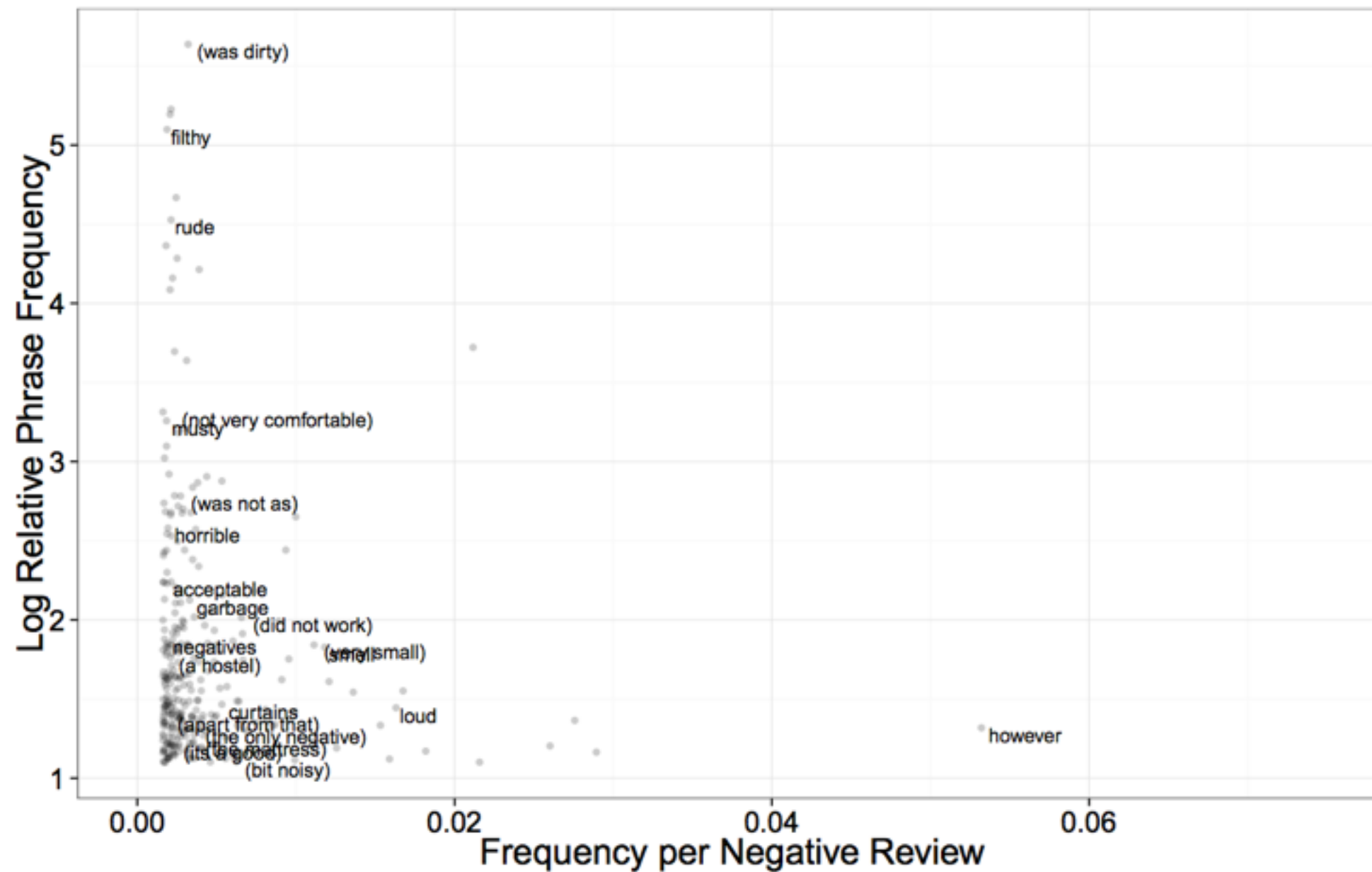
[illegible]



yawitzd.github.io/airbnb



Figure A1: Distribution of negative phrases in guest reviews of listings.



“Relative phrase frequency” refers to the ratio with which the phrase occurs in reviews with a rating of less than five stars.

[http://andreyfradkin.com/assets/reviews\\_paper.pdf](http://andreyfradkin.com/assets/reviews_paper.pdf)

```
In [56]: bad = r[r["comments"].str.contains("rude|dirty|unfortunately")]
```

```
        I would not 111
        I would n't 108
    I would recommend 98
would not recommend 94
    I would definitely 88
unfortunately n't get 86
        n't get meet 81
        made us feel 76
    The apartment clean 69
    The location great 68
        I could n't 61
would definitely stay 58
        I ca n't 56
        I 'm not 54
within walking distance 52
        I would say 50
    The apartment well 50
        I n't think 46
        I n't know 43
    would n't recommend 42
apartment great location 41
    The apartment not 39
```

```
        I n't get 39
        I would stay 39
        New York City 38
would definitely recommend 38
apartment well located 38
        I could not 38
    The location good 37
SENSITIVE CONTENTS HIDDEN 36
    The room clean 36
    little bit dirty 36
    The room small 36
    spend much time 35
    The bathroom dirty 35
not recommend place 35
    great place stay 35
    us feel welcome 35
    I unfortunately n't 34
    The location apartment 34
```

# LDA - Topic Modeling from all 400,000 reviews

Topic #0: CONVENIENT LOCATION

located, well, perfectly, well located, rooms, nyc, conveniently, far, soho, conveniently located, anywhere

Topic #1: WALKING DISTANCE

walk, walking, subway, apartment, us, minutes, get, distance, within, walking distance, easy, manhattan,

Topic #2: PLACE AS NEEDED

nice, really, everything, place, needed, described, exactly, everything needed, recommended, stay, host,

Topic #3: PLACE LIKE HOME

back, come, home, amazing, apt, come back, could, better, best, large, perfect, nyc, couldn, spent, not,

Topic #4: NEW YORK

new, york, new york, und, enjoyed, die, really enjoyed, enjoyed stay, advertised, city, apartment, york

Topic #5: NICE LOCATION

great, location, apartment, great location, excellent, stay, clean, host, apartment great, great stay, s

Topic #6: NEGATIVE

not, apartment, us, didn, room, good, get, would, time, one, night, place, even, could, day, little, use

Topic #7: GREAT HOST

great, definitely, host, stay, great host, would, clean, would definitely, apartment, definitely stay, h

Topic #8: THINGS TO DO

restaurants, apartment, bars, check, plenty, shops, east, village, side, close, restaurants bars, cosy, r