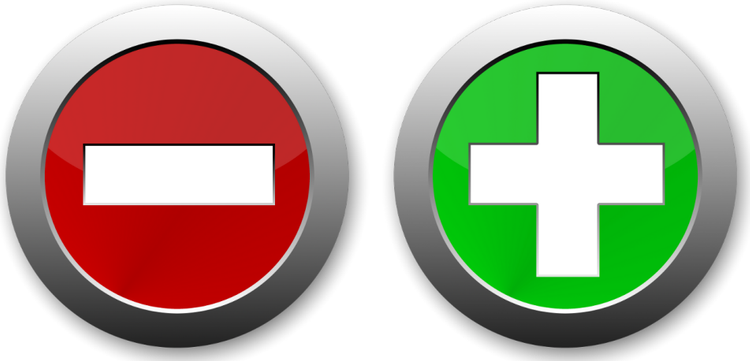
Let's talk about our feelings

SENTIMENT ANALYSIS IN R

Ted Kwartler
Data Dude

Definition: sentiment analysis

Sentiment analysis is *the process of extracting an author's emotional intent from text*



Why is sentiment analysis important?

Data formats in this course

Bag of Words DTM & TDM

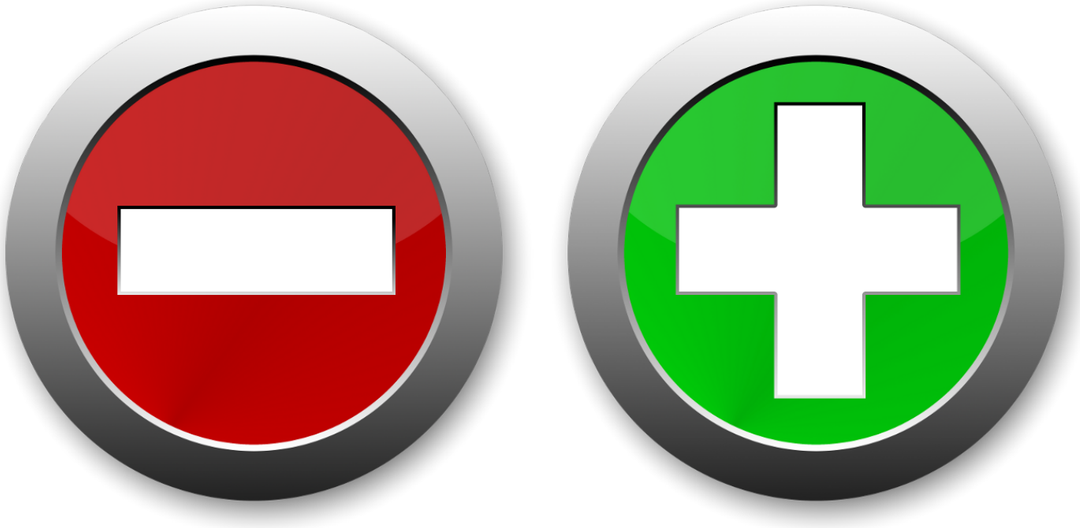
Docs	yeah	yeahah	yeahand	yeahgod	yeahhh	yeahho	yeahlong
1	8	0	0	0	0	0	0
2	1	0	0	0	0	0	0

Tidy Tribble...*errr*...Tibble

```
> tidy.rappers
Source: local data frame [1,525,121 x 6]
Groups: artist [12]

tidy.rappers[,3:6]
                          song_title  word original_word_order artist_song_id
                               <chr> <chr>                <int>          <int>
1  187 Um (deep Cover Remix) lyrics  yeah                    1              1
2  187 Um (deep Cover Remix) lyrics   and                    2              1
3  187 Um (deep Cover Remix) lyrics   you                    3              1
4  187 Um (deep Cover Remix) lyrics don't                    4              1
5  187 Um (deep Cover Remix) lyrics  stop                    5              1
6  187 Um (deep Cover Remix) lyrics  yeah                    6              1
7  187 Um (deep Cover Remix) lyrics   and                    7              1
8  187 Um (deep Cover Remix) lyrics   you                    8              1
9  187 Um (deep Cover Remix) lyrics don't                    9              1
10 187 Um (deep Cover Remix) lyrics  stop                   10              1
```

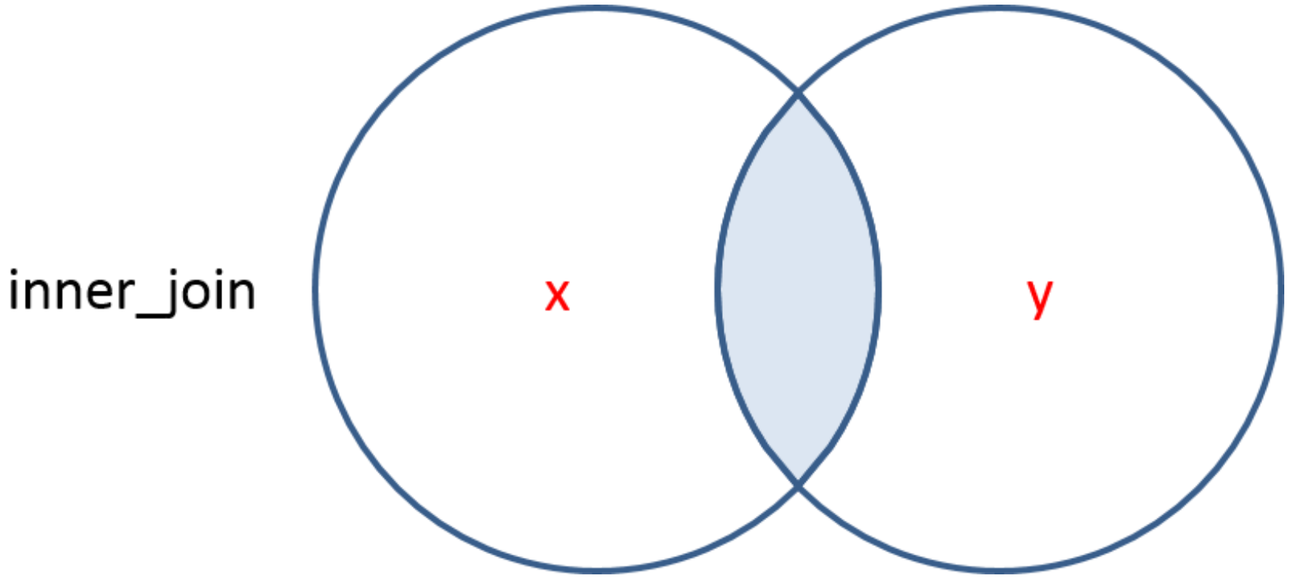
Chapter 1: qdap's polarity() function

```
library(qdap)

polarity(text$column)

polarity(text$column, text$factor_or_author_grouping)
```



Chapter 2: tidytext inner joins

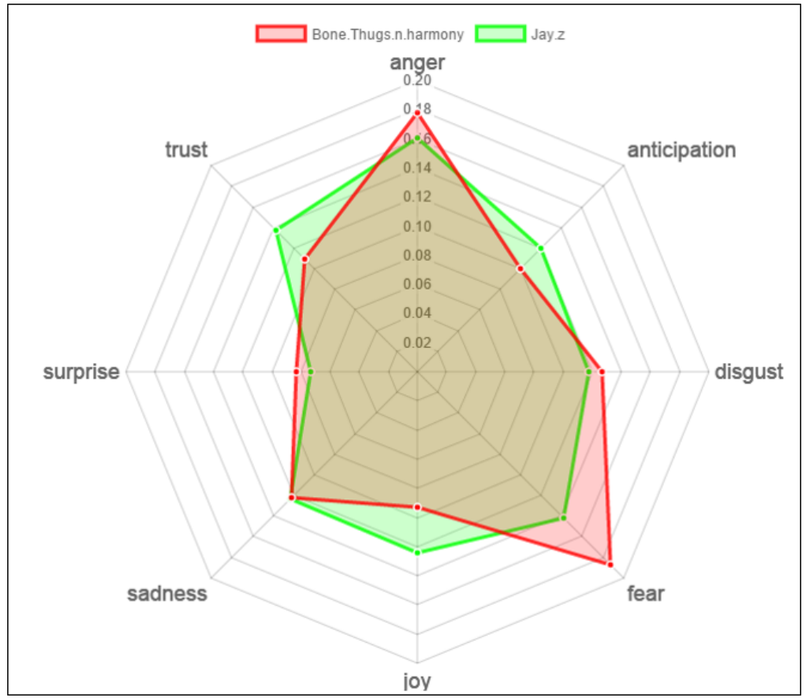
```
library(tidytext)


inner_join(sentiment_words, some_text_to_be_analyzed)
```

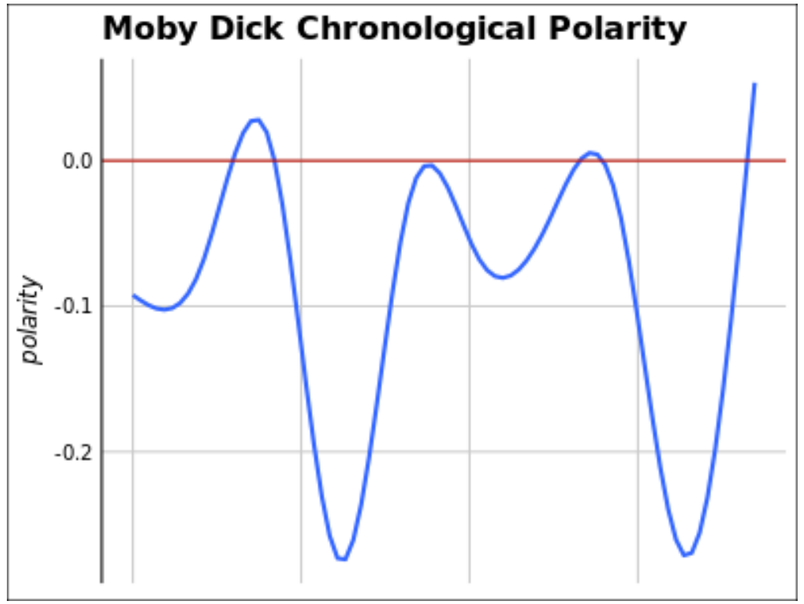


Chapter 3: Visualizing sentiment

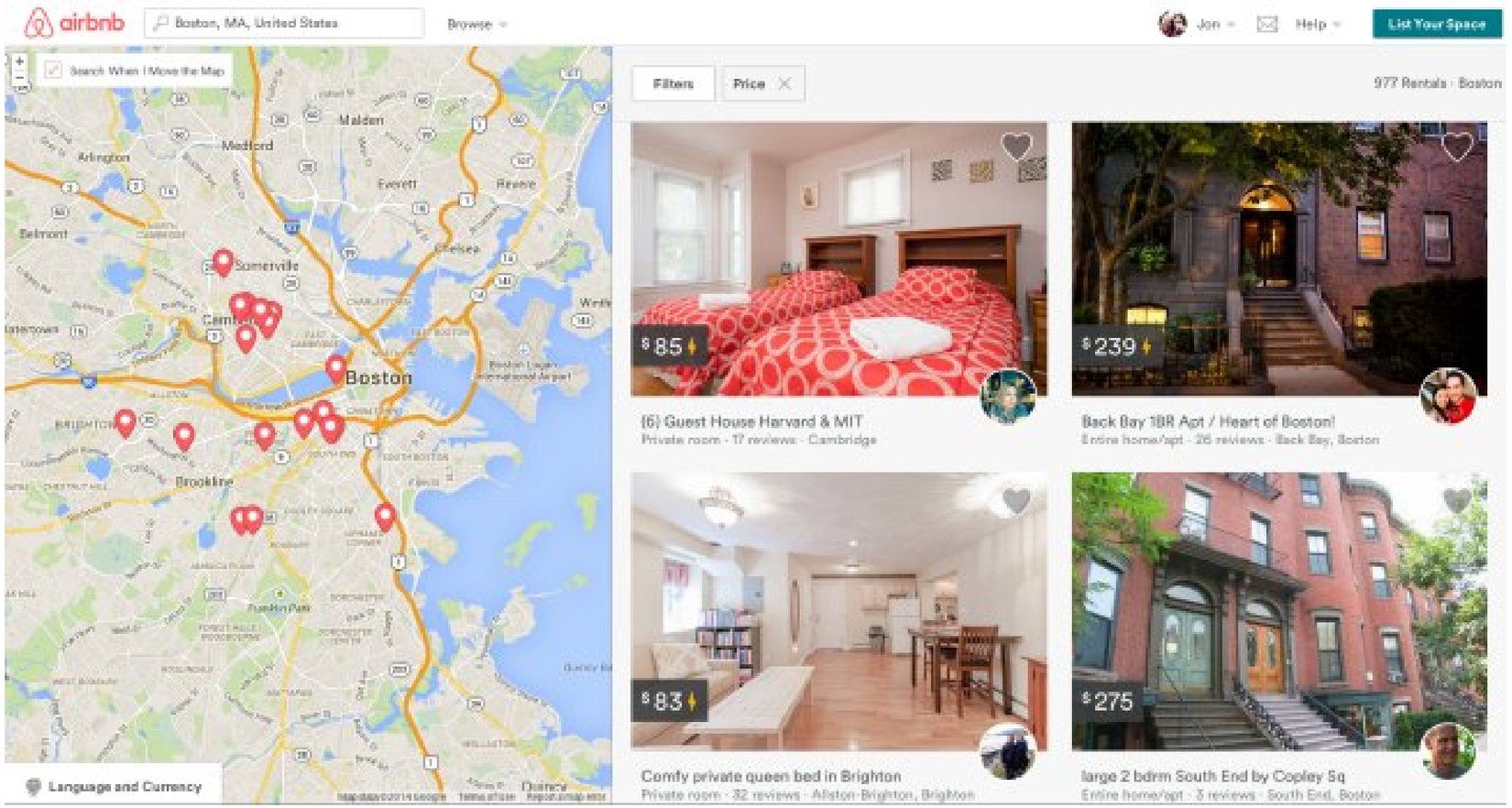
htmlwidgets.org radar chart

`ggplot2` line chart





Chapter 4: Case study on property rentals



Let's practice!

SENTIMENT ANALYSIS IN R

How many words do YOU know? Subjectivity lexicons, Zipf's Law & Least Effort

SENTIMENT ANALYSIS IN R

Ted Kwartler
Data Dude

Subjectivity lexicon

```
library(qdap)
library(magrittr)

text_df %$% polarity(text)
```

Returns a "polarity" object with positive and negative scores.

A **subjectivity lexicon** is a predefined list of words associated with emotional context such as positive/negative, or specific emotions like "frustration" or "joy."

Where to get subjectivity lexicons?

- `qdap` 's `polarity()` function uses a lexicon from `hash_sentiment_huliu`
- `tidytext` has a `sentiments` tibble with
  - **NRC** - Words according to 8 emotions like "angry" or "joy" and Pos/Neg
  - **Bing** - Words labeled positive or negative
  - **AFINN** - Words scored from -5 to 5

library(lexicon)

Name	Description
`dodds_sentiment`	Mechanical Turk Sentiment Words
`hash_emoticons`	Translations of basic punctuation emoticons :)
`hash_sentiment_huliu`	U of IL @CHI Polarity (+/-) word research
`hash_sentiment_jockers`	A lexicon inherited from `library(syuzhet)`
`hash_sentiment_nrc`	5468 words crowdsourced scoring between -1 & 1

No way! Too few words.

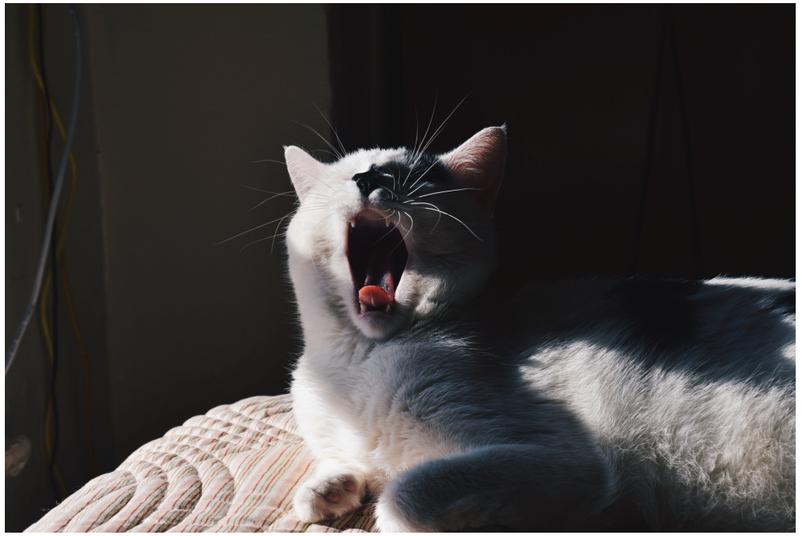
- Zipf's Law
- Principle of Least Effort

Zipf's Law in action

Rank	City	2010 Census Population	Actual %	Zipf's Expected %
1	New York	8,175,133	100%	...
2	LA	3,792,621	46%	50%
3	Chicago	2,695,598	33%	33%
4	Houston	2,100,263	26%	25%
5	Philadelphia	1,526,006	19%	20%

Principle of Least Effort

If there are several ways of achieving the same goal, people will choose the least demanding course of action



Up next...

Let's practice!

SENTIMENT ANALYSIS IN R

Explore qdap's polarity & built-in lexicon

SENTIMENT ANALYSIS IN R

Ted Kwartler
Data Dude

polarity()

An example subjectivity lexicon:

Word	Polarity
Amazing	Positive
Bad	Negative
Good	Positive
...	...
Wonderful	Positive

Context cluster

Example context cluster:

The DataCamp sentiment course is very GOOD for learning.

Context cluster, continued

Example context cluster:

The DataCamp sentiment course is very GOOD for learning.

Term	Class	Word Count
Very	Amplifier	1
Good	Polarized Term/Positive	1
All other words	Neutral	7

Context cluster glossary

- **Polarized Term** - words associated with positive/negative
- **Neutral Term** - no emotional context
- **Negator** - words that invert polarized meaning e.g. "not good"
- **Valence Shifters** - words that effect the emotional context
  - **Amplifiers** - words that increase emotional intent
  - **De-Amplifiers** - words that decrease emotional intent

Context cluster scoring

Example context cluster:

The DataCamp sentiment course is very GOOD for learning.

Term	Class	Word Count	Polarity Value
Very	Amplifier	1	0.8
Good	Polarized Term/Positive	1	1
All other words	Neutral	7	0

Polarity calculation

Class	Word Count	Polarity Value
Amplifier	1	0.8
Polarized Term	1	1
Neutral	7	0
Sum	**9**	**1.8**

Example Context Cluster

The DataCamp sentiment course is very GOOD for learning.

1. 1 + 0.8 = 1.8
2. 1+1+7 = 9
3. $$\frac{1.8}{\sqrt{9}}$$

Answer: 0.6

Let's practice!

SENTIMENT ANALYSIS IN R