

# **Basic Text Mining and Sentiment Analysis**

**Presenter: Wenxiao Zhou**

**Department of Statistics, University of Connecticut**



# Inspiration



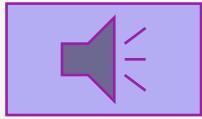
**Once you are listening to a music, or watching a movie:**

*Happy or Sad? Love or Hate?  
Hopeful or Despair?  
Passionate or Indifferent?*

## **Can any of the emojis below represent your emotion?**



# Listen to a song: The Cure -- Lady Gaga



*Music in general is positive,  
encouraging and healing.*

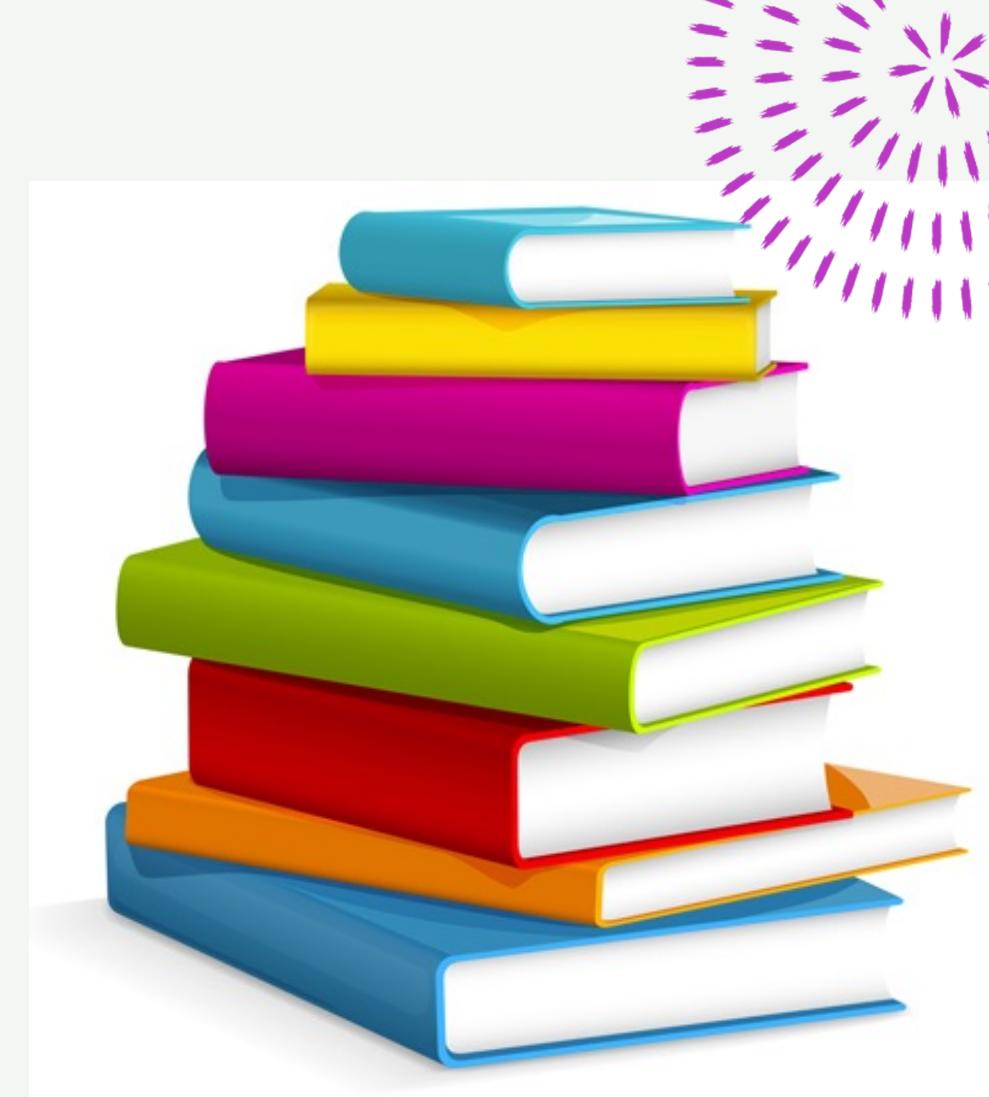
*Question: How can computer  
detect emotions like human???*

If I can't find the cure, I'll  
I'll fix you with my love  
No matter what you know, I'll  
I'll fix you with my love  
And if you say you're okay  
I'm gonna heal you anyway  
Promise I'll always be there  
Promise I'll be the cure (be the cure)

# **Sentiment Analysis:**

## **A process extracting an author's emotional intent from text.**

**It is a natural language processing (NLP) technique used to determine whether data is positive, negative or neutral.**



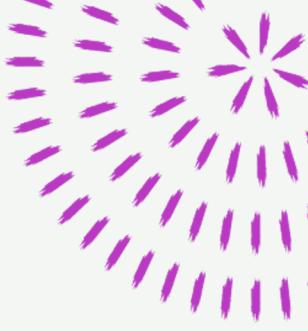
# Why is Sentiment Analysis Important?

- There are many apps setting comments platforms for their customers to make reviews, sentiment analysis helps to monitor and automatically analyzing customer feedback.



## Recall our Question:

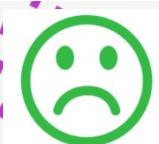
## How can computer detect emotions like human???



1. Detect the texts and break it into tidy format components



2. Make comparison with lexicons, assign sentiment score



3. Combine scores and data visualization



# Fast & Dirty Analysis: qdap library, polarity function

polarity():

Use an academic lexicon from the University of Illinois-Chicago.  
It contains almost 7000 words marked as positive or negative.

key.pol

#Negators

negation.words

#Amplifiers

Amplifiers

amplification.words

# De-amplifiers

deamplification.words

- 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



# Note:

- Remember to download Java to support qdap package first, find the version for your computer in the link below:
- <https://www.oracle.com/java/technologies/javase-jdk15-downloads.html>
- To check you can use qdap package by the following codes:

```
install.package("qdap")
Install.package("rJava")
library(qdap)
library(rJava)
help(polarity)
```

# Example

Titanic is a **very** **great** movie.

Term	Class	Word Count	Polarity Value
Very	Amplifier	1	0.8
great	Polarized Term/Positive	1	1
All other words	Neutral	5	0

Polarity calculation:

$$1+0.8=1.8$$

$$1+1+5=7$$

Polarity

$$\text{score}=1.8/\sqrt{7}=0.6803$$

## **Question:**

# **How can subjectivity lexicons handle various types of human emotions ?**

**Two principals:**

**1. Zipf' Law**

**2. The Principle of least efforts**



Even though two principals can make it logically use subjectivity lexicons, you will still need to adjust lexicons to fit the text source under certain environments, e.g. different age groups may have special words represents emotions.

# Have a Try!!!

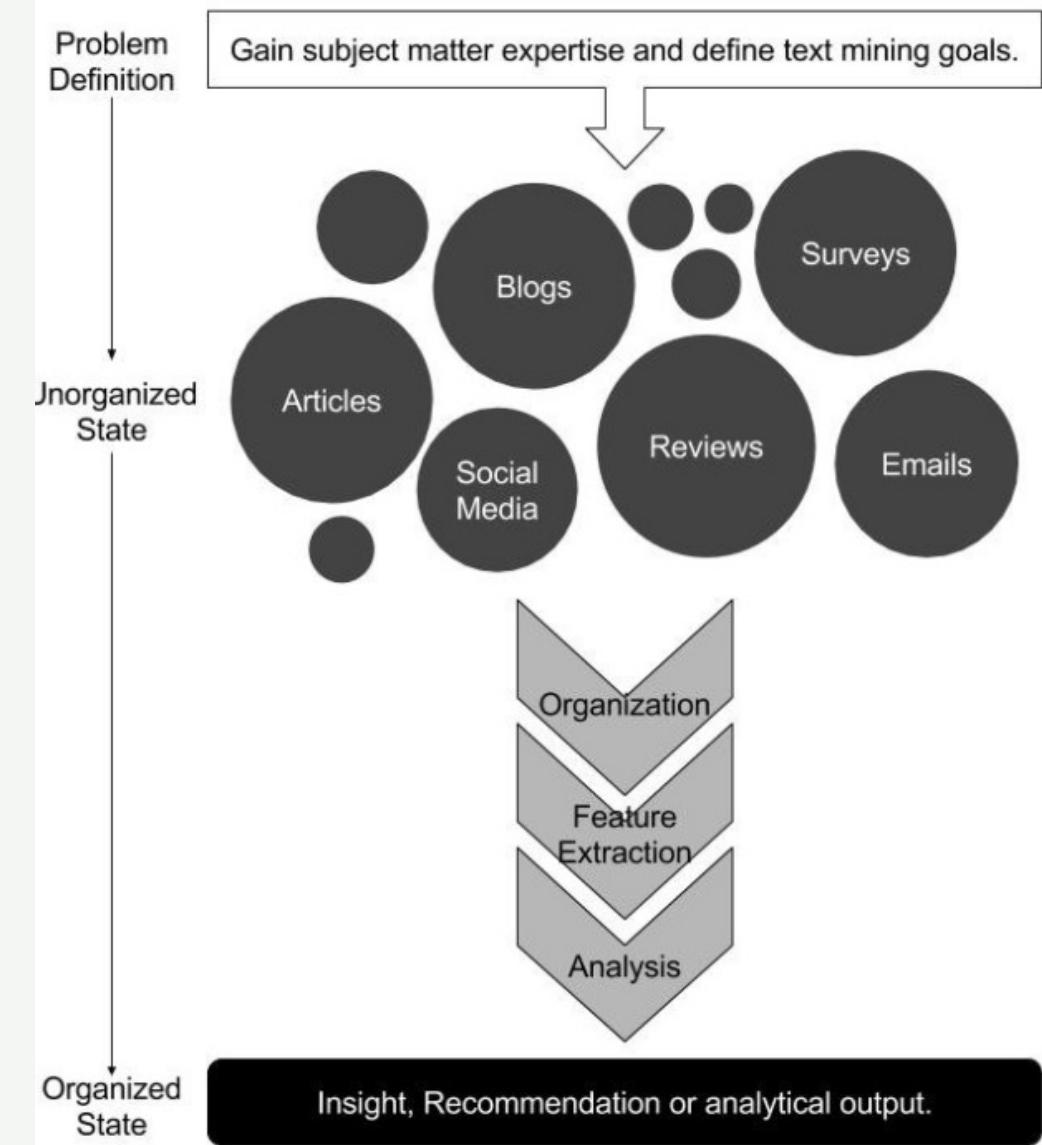
- Let's go back to our **Lyrics Analysis on "The Cure"**
- A fast & Dirty **Sentiment Analysis in R**

# More interesting!!!

- Advanced lyrics analysis on **Taylor Swifts songs**

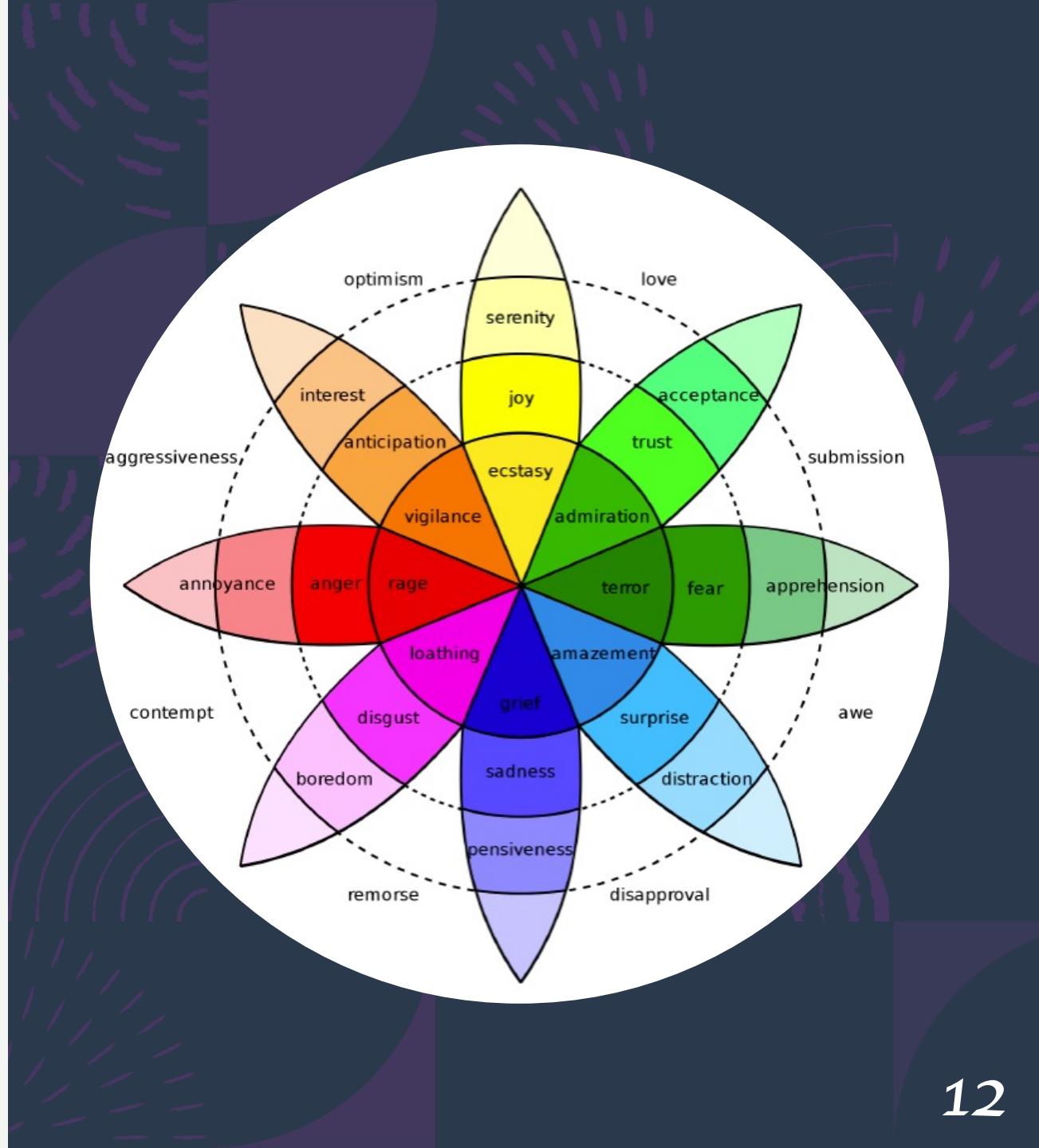
**Go Further:  
What if we are interested  
in analyzing the emotions  
of customers' comments in  
one platform?**

## Text Mining Workflow



# Plutchik's Wheel of Emotion

- Visualization consisting of eight primary emotions including joy, trust, fear, surprise, sadness, disgust, anger, anticipation.



# Common Subjectivity Lexicons & Inner Join

Bing Lexicon & inner  
join

NRC Lexicon

AFINN Lexicon

# inner join

## **dplyr package** join

```
inner_join(x,y,...)  
left_join(x,y,...)  
right_join(x,y,...)  
semi_join(x,y,...)  
anti_join(x,y,...)
```

## **declaring the by parameter:**

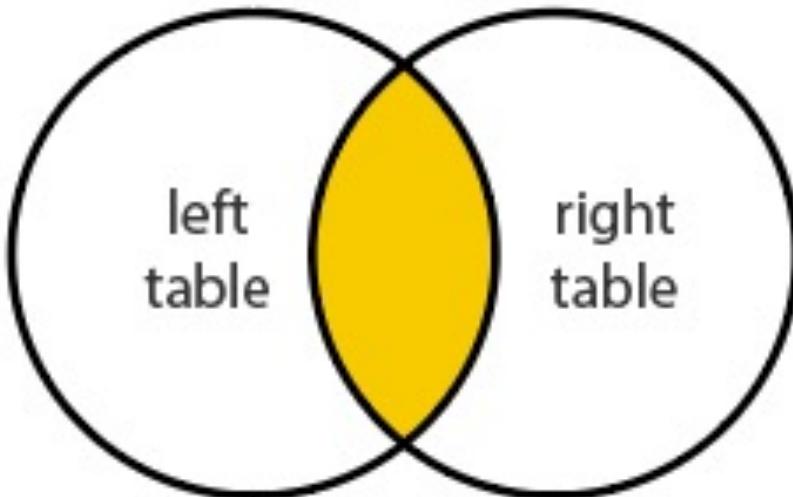
```
inner_join(x,y,by="shared_column")
```

**or**

```
inner_join(x,y,by=c("a"="b"))
```

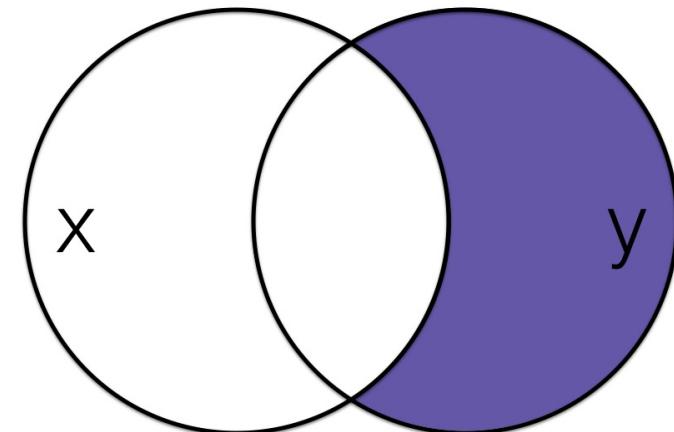
# Comparing inner join and anti joins

INNER JOIN



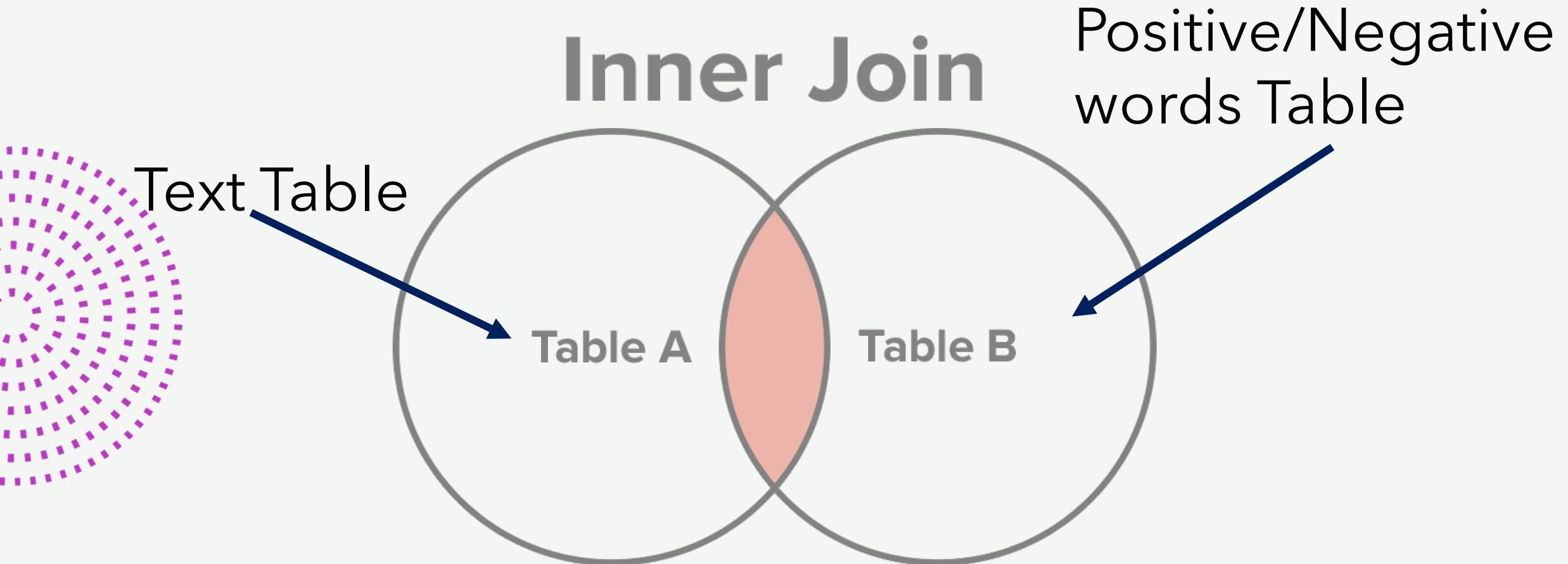
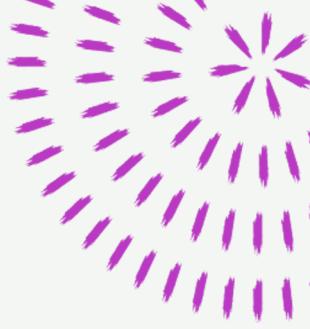
```
inner_join(  
  text_table,  
  subjectivity_lexicon,  
  by="word_column"  
)
```

anti\_join(y, x)



```
anti_join(  
  text_table,  
  stopwords_table,  
  by="word_column"  
)
```

In our case  
**Inner join Texts with positive/negative words**



# Basic Introduction:

The **NRC** lexicon categorizes words in a binary fashion ("yes"/"no") into categories of positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust.

The **BING** lexicon categorizes words in a binary fashion into positive and negative categories.

The **AFINN** lexicon assigns words with a score that runs between -5 and 5, with negative scores indicating negative sentiment and positive scores indicating positive sentiment.

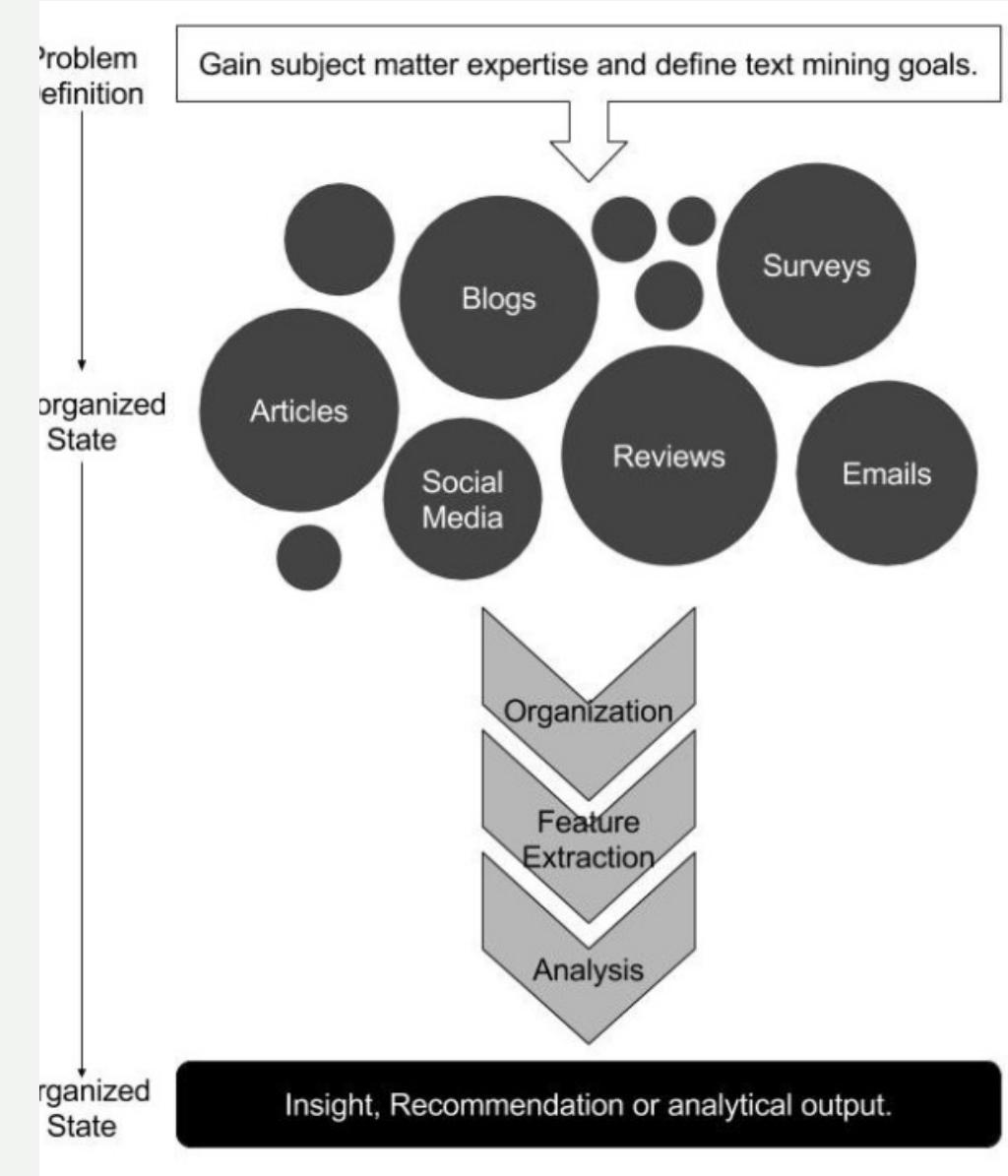


```
library(tidytext)  
library(textdata)  
get_sentiments("afinn")  
get_sentiments("nrc")  
get_sentiments("bing")
```

# Recall: Text Mining Workflow

## Six Steps:

- Define the Problem & Specific Goals
- Identify the text
- Organize the text
- Extract features
- Analyze
- Draw a conclusion/ reach an insight



# Have a Try!!!

- Let's go to our advanced case study
- Airbnb Reviews Sentiment Analysis in R

# Summary

- qdap package, polarity() function
- Text Mining workflow & Basic Text cleaning
- Subjectivity lexicons & Inner join
- Two Cases Study
  - Data and Codes download:  
<https://github.com/WenxiaoZhou/My-projects/tree/main/Text%20Mining%20and%20Sentiment%20Analysis>



# References

1. KWARTLER, E. (Ed.). (n.d.). Text Mining with Bag-of-Words in R. (website)  
<https://learn.datacamp.com/courses/intro-to-text-mining-bag-of-words>
2. Tatman, R. (2017, September 15). Tutorial: Sentiment analysis in r. (website)  
<https://www.kaggle.com/rtatman/tutorial-sentiment-analysis-in-r>
3. Silge, J., & Robinson, D. (2017). Text mining with R: A tidy approach. Sebastopol, CA: O'Reilly  
<https://www.tidytextmining.com/index.html>

# Thank you for listening!!!

Welcome to contact me:

Email: [wenxiao.zhou@uconn.edu](mailto:wenxiao.zhou@uconn.edu)

Linkedin: [linkedin.com/in/wenxiao-christine-zhou](https://linkedin.com/in/wenxiao-christine-zhou)