

A close-up portrait of a young woman with dark hair pulled back, wearing large brown-rimmed glasses and a light blue top. She is looking directly at the camera with a slight smile. The background is a solid blue.

May 2020

# CONVERSATION ANALYTICS

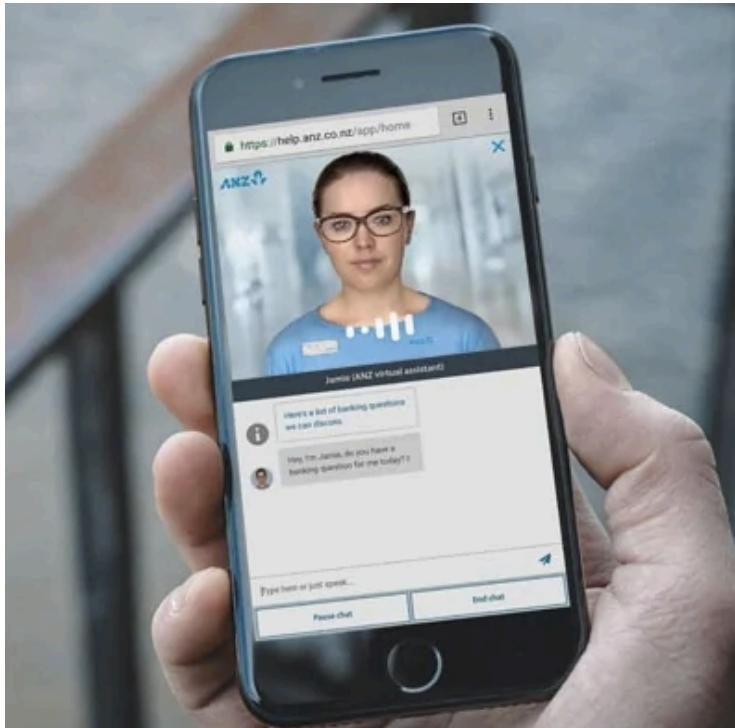
## Understanding a digital assistant with R

---

Jayne Ihaka, ANZ New Zealand  
@jayniehaka



## ABOUT ME



- Studied biotechnology and finance
- Worked at ANZ for 7 years
- Developed interest in data science through online courses, meetups, conferences
- Currently working on a digital assistant, "Jamie"

## QUESTIONS TO BE ANSWERED



What are users asking Jamie about?



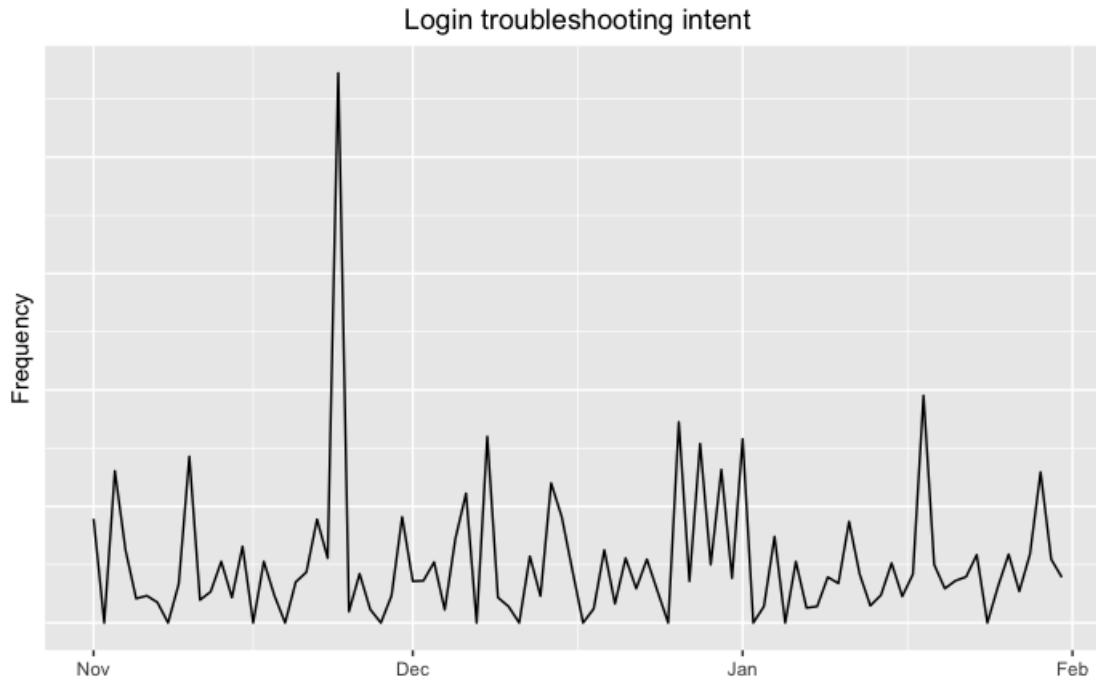
How successful are these conversations?



What can we do to make it better?

## WHAT ARE CUSTOMERS ASKING JAMIE ABOUT?

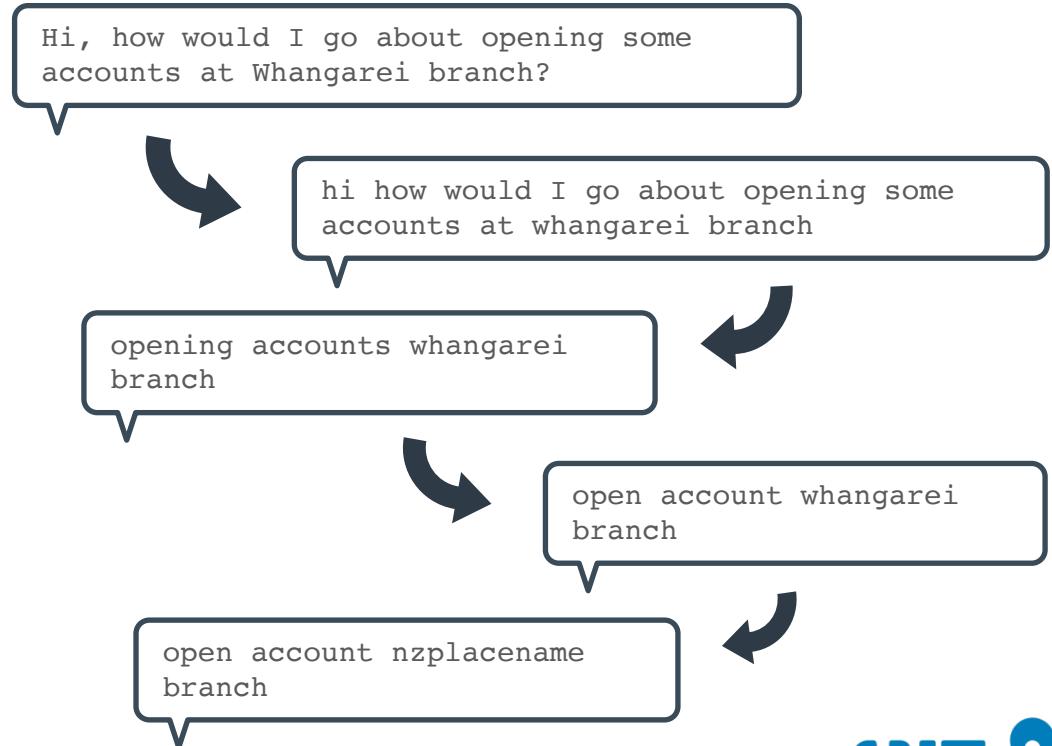
Look at top intents and frequency over time:



# PRE-PROCESSING CONVERSATION TRANSCRIPT DATA

- Remove everything but user utterances
- Strip numbers, special characters, & upper case
- Replace contractions
- Stop words
- Stemming/lemmatising
- Replacements and grouping

**R packages:**  
tidytext  
stopwords  
textclean  
textstem



## CUSTOM REPLACEMENTS AND GROUPING

log in      login  
log on      logon

singapore      london  
australia      fiji

visa      cc  
credit card



login



overseas placename



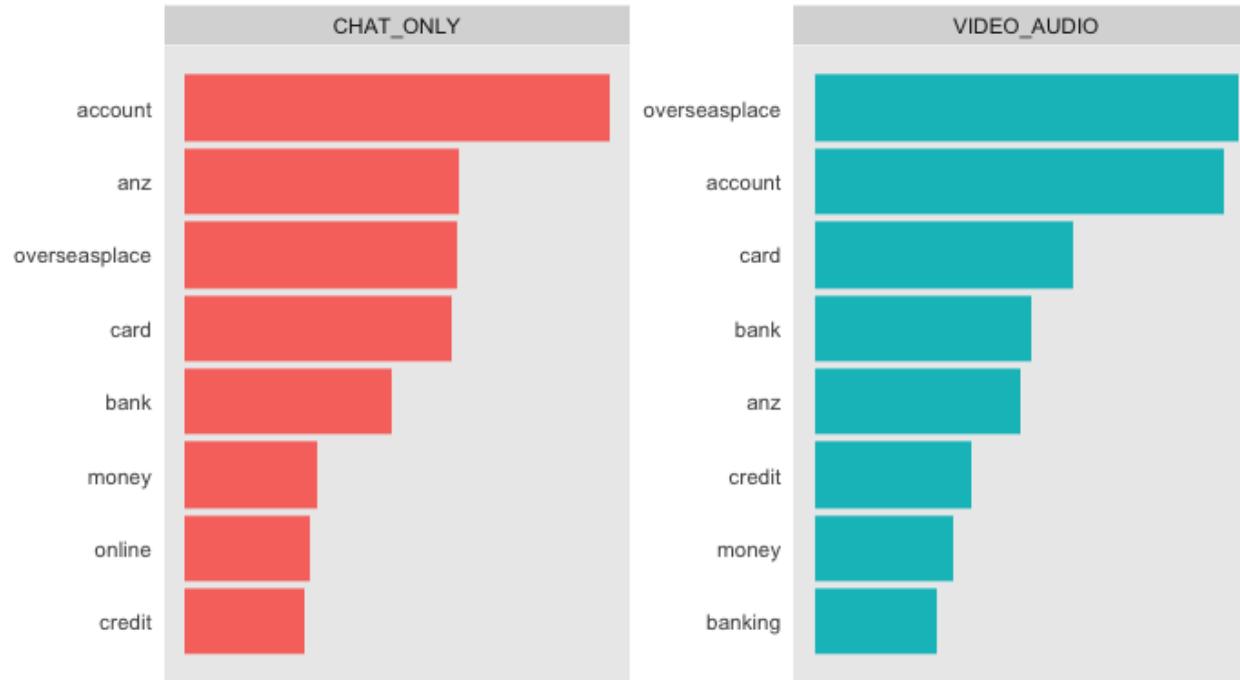
credit card

## PRE-PROCESSING TIPS

- Think carefully about the order of these steps!
- Pick the right libraries and word lists for your use case
- Account for domain-specific jargon, common misspellings, slang, txt speak
- Know that you will lose some information at each step
- Trade-off between accuracy and effort

# WHAT ARE USERS ASKING ABOUT?

Simply comparing popular words is not very useful:

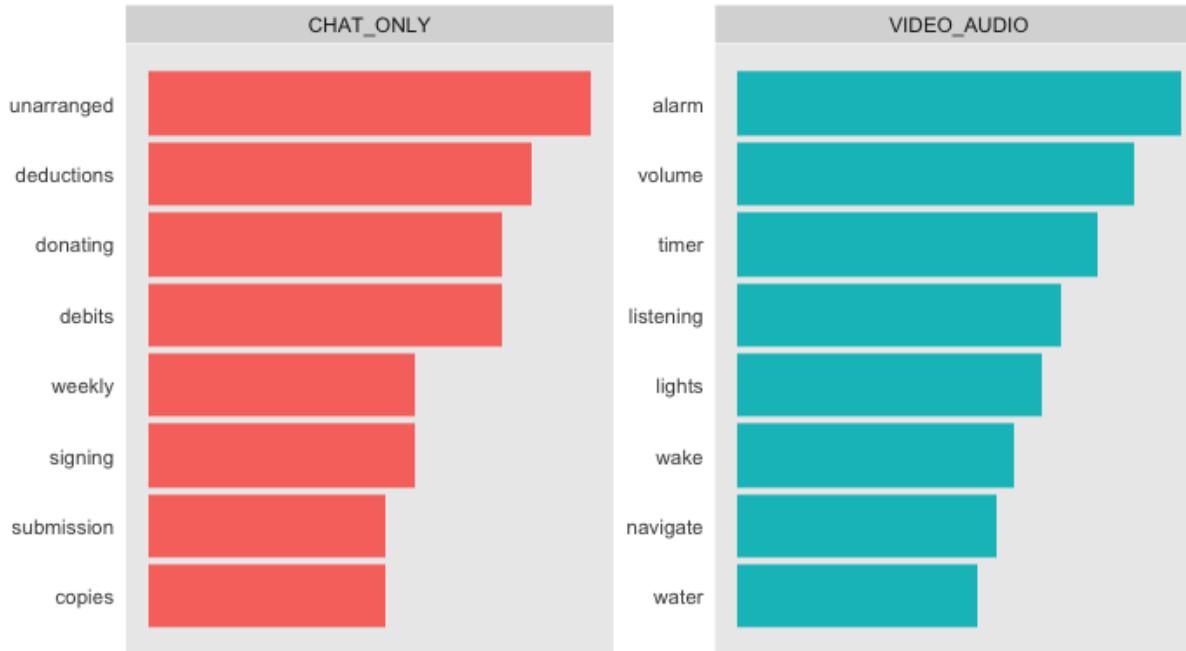


R packages:  
tidytext



## WHAT ARE USERS ASKING ABOUT?

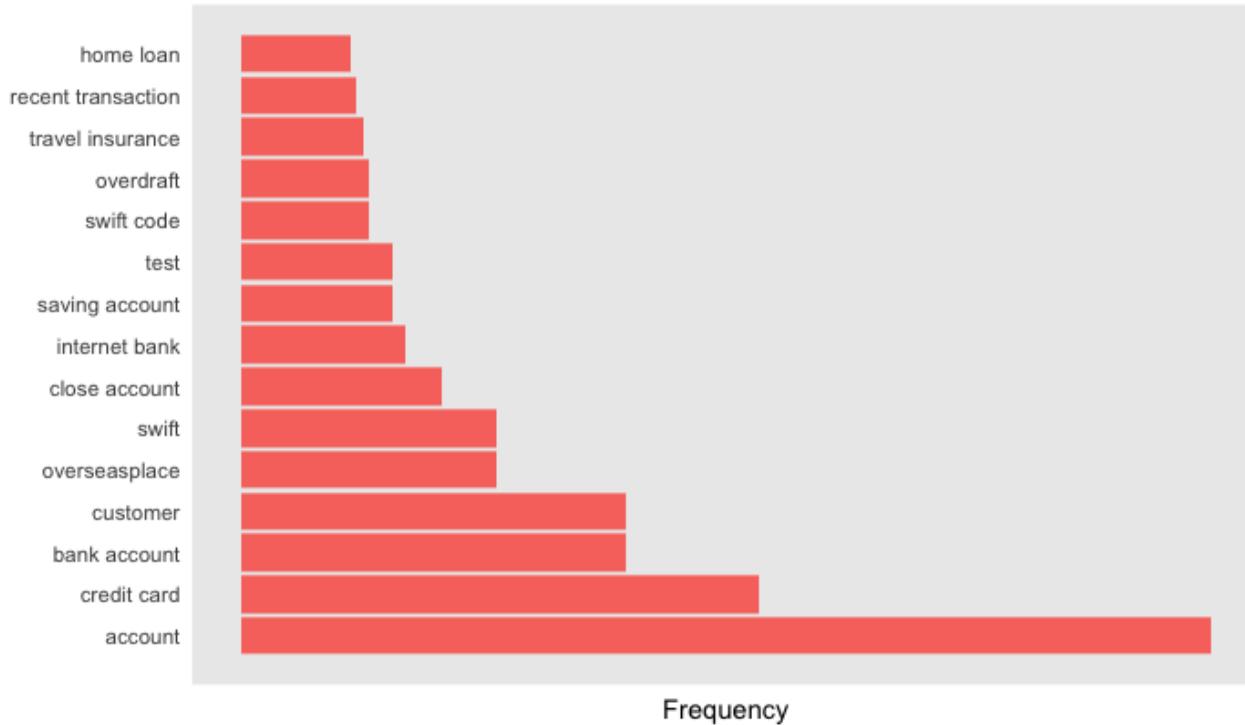
Popular words using term frequency-inverse document frequency (TD-IDF):



R packages:  
tidytext  
tm

# WHAT ARE USERS ASKING ABOUT?

Rapid Automatic Keyword Extraction (RAKE):



**R packages:**  
udpipe



# WHAT ARE USERS ASKING ABOUT?

Topic modelling using Latent Dirichlet Allocation (LDA) method:

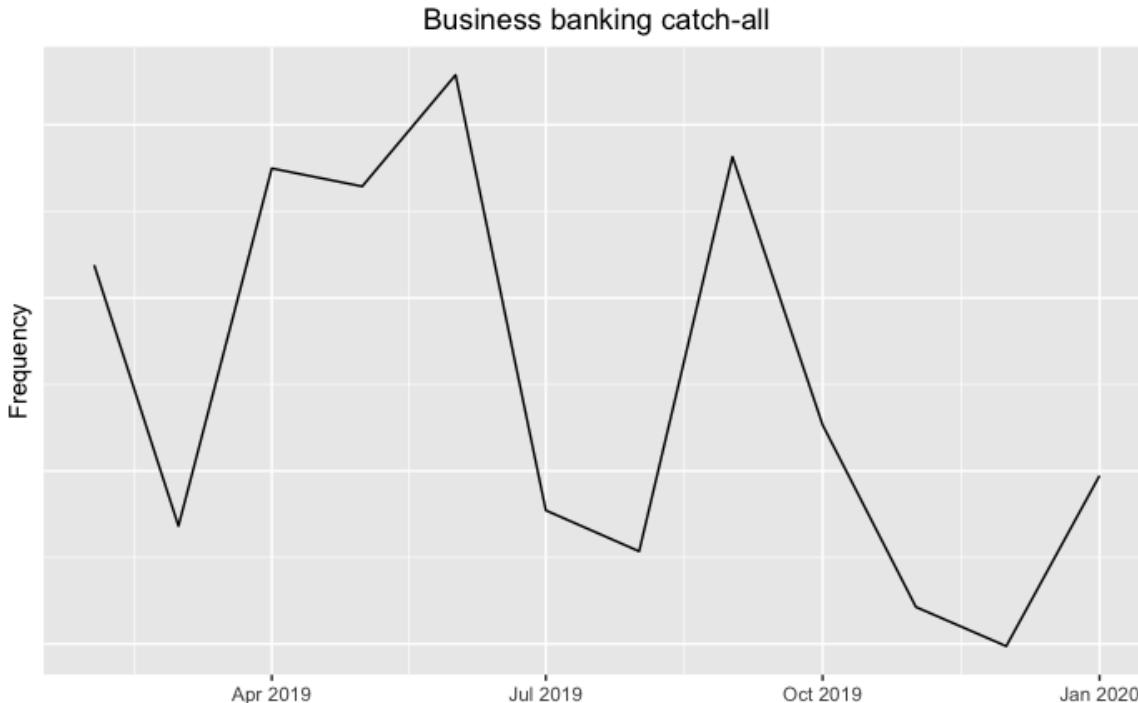
Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
money	bank	card	overseasplace	overseasplace	anz	joke	account
overseas	online	credit	time	loan	branch	play	bank
transfer	customer	visa	question	home	overseasplace	favourite	anz
overseasplace	internet	debit	talk	rate	email	start	nzplacename
payment	phone	travel	bye	deposit	address	weather	close
send	change	cancel	day	apply	call	people	set
nzplacename	password	lose	answer	fix	contact	google	id
code	forget	insurance	person	kiwisaver	personname	christmas	check
anz	log	transaction	hear	mortgage	appointment	pretty	saving
pay	app	block	speak	overdraft	book	song	live

R packages:  
topicmodels



# WHAT ARE CUSTOMERS ASKING JAMIE ABOUT?

Track a new topic using elegant failures:

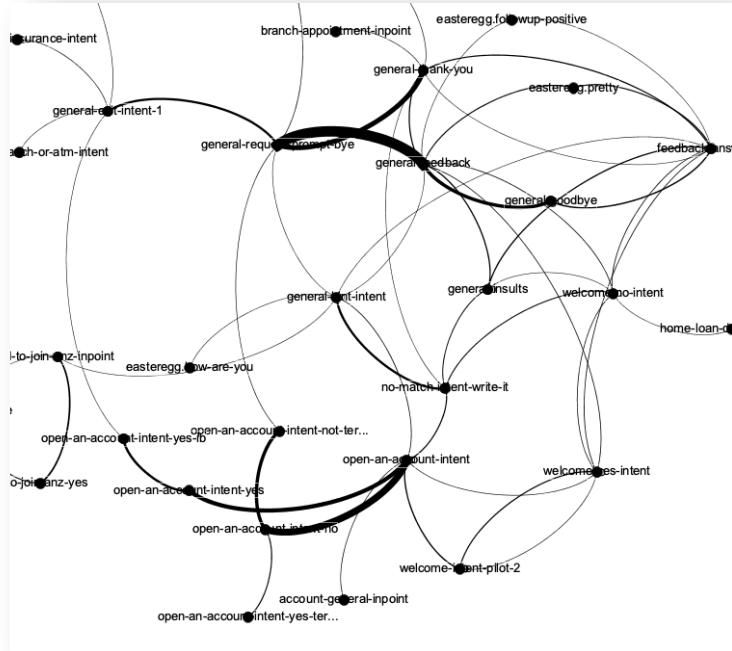


# HOW DO CONVERSATIONS FLOW?

Network analysis:

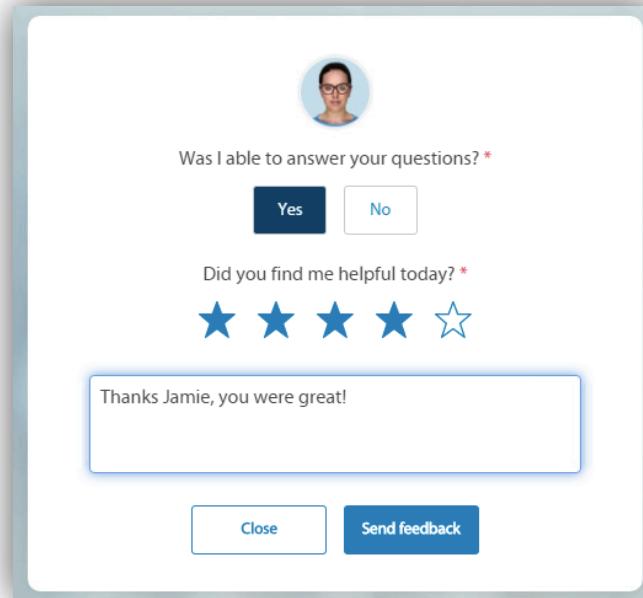
- Conversations can be modelled as a directed graph
- Each node is an intent, links are utterances
- See emerging pathways
- See common issues such as loops

R packages:  
igraph  
ggnet



## HOW SUCCESSFUL ARE THE CONVERSATIONS?

- Ask the user: star rating, CSAT, NPS
- Confidence scores
- Sentiment analysis
- Define points in the conversation as successful
- Get a human to read it!



## SUMMARY

- R is awesome for extracting insights from conversations
- Pre-processing is crucial and needs to be tailored to your particular use case
- Measuring success in an automated way is hard
- Analytics can help - but it still needs humans to dig deeper and validate

# Q & A

## HELPFUL RESOURCES

Presentation on text mining by Jared Lander

<https://jaredlander.com/content/2019/03/TextMining#1>

Text Mining with R - Julia Silge and David Robinson

<https://www.tidytextmining.com/>

CRAN Task View: Natural Language Processing

<https://cran.r-project.org/web/views/NaturalLanguageProcessing.html>

Feedback analysis - LinkedIn post by Alyona Medelyan

<https://www.linkedin.com/pulse/how-build-your-own-feedback-analysis-solution-alyona-medelyan-phd/>

