



EXTRACTIVE AND ABSTRACTIVE TEXT SUMMARIZATION OF SOUTH PARK EPISODES

By: Tim Rabbitt



OVERVIEW

- A new website wants to **use character dialogue** in popular TV shows to create **episode summaries**
- Starting with **South Park**

THE DATA

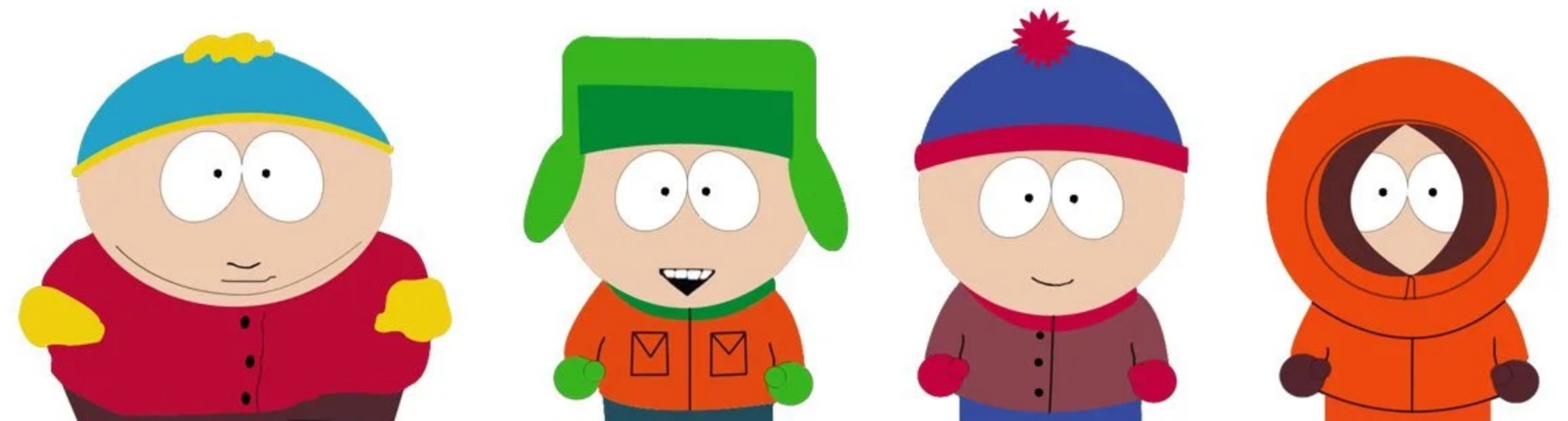
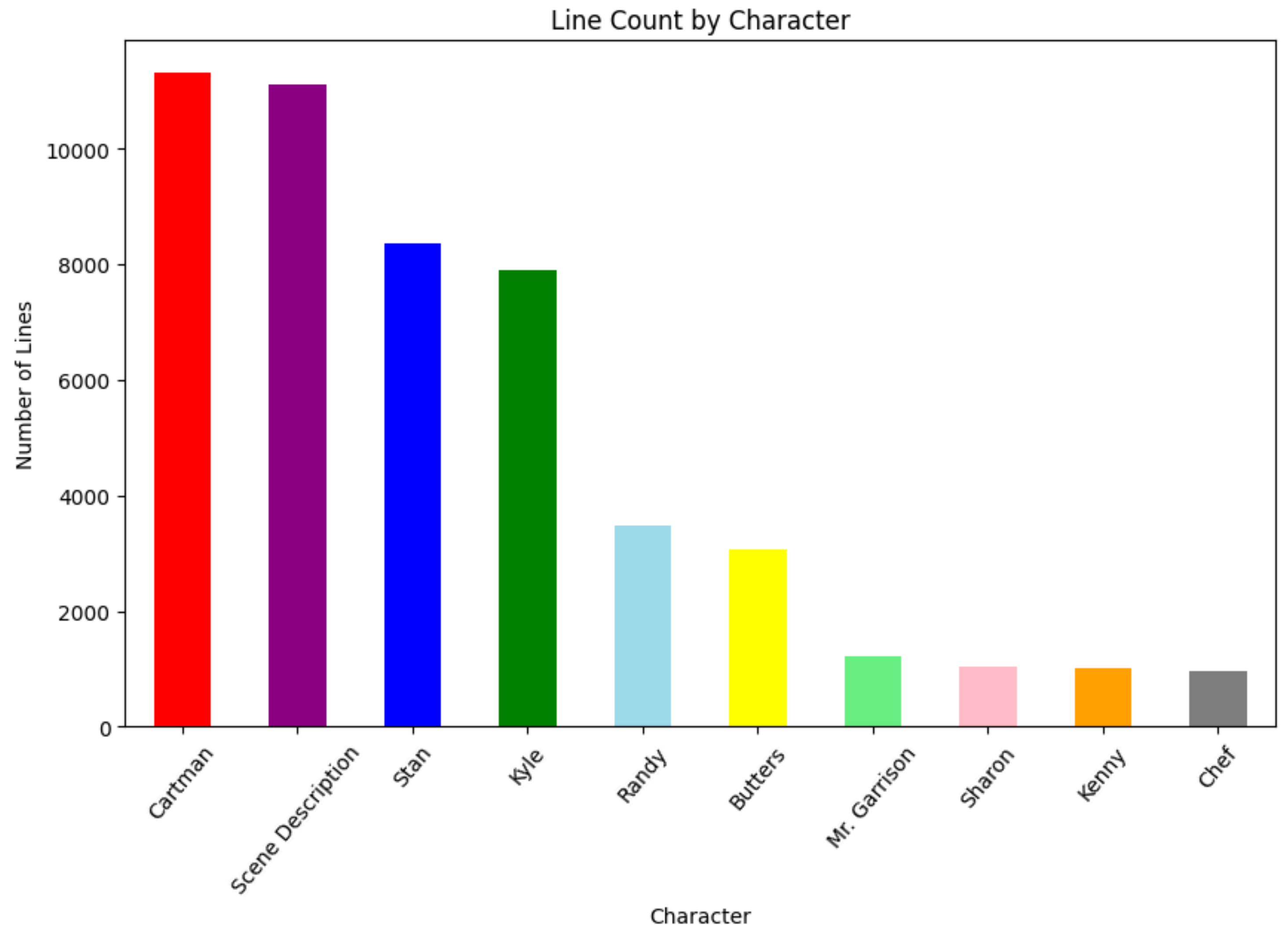


Kaggle South Park Scripts Dataset

- **300** episodes
- **95,000** lines
- Episode name, **description**, and **character lines**

DATA EXPLORATION

- Cartman, Stan, and Kyle contribute the most to the dialogue
- Scene description provides setting and character actions



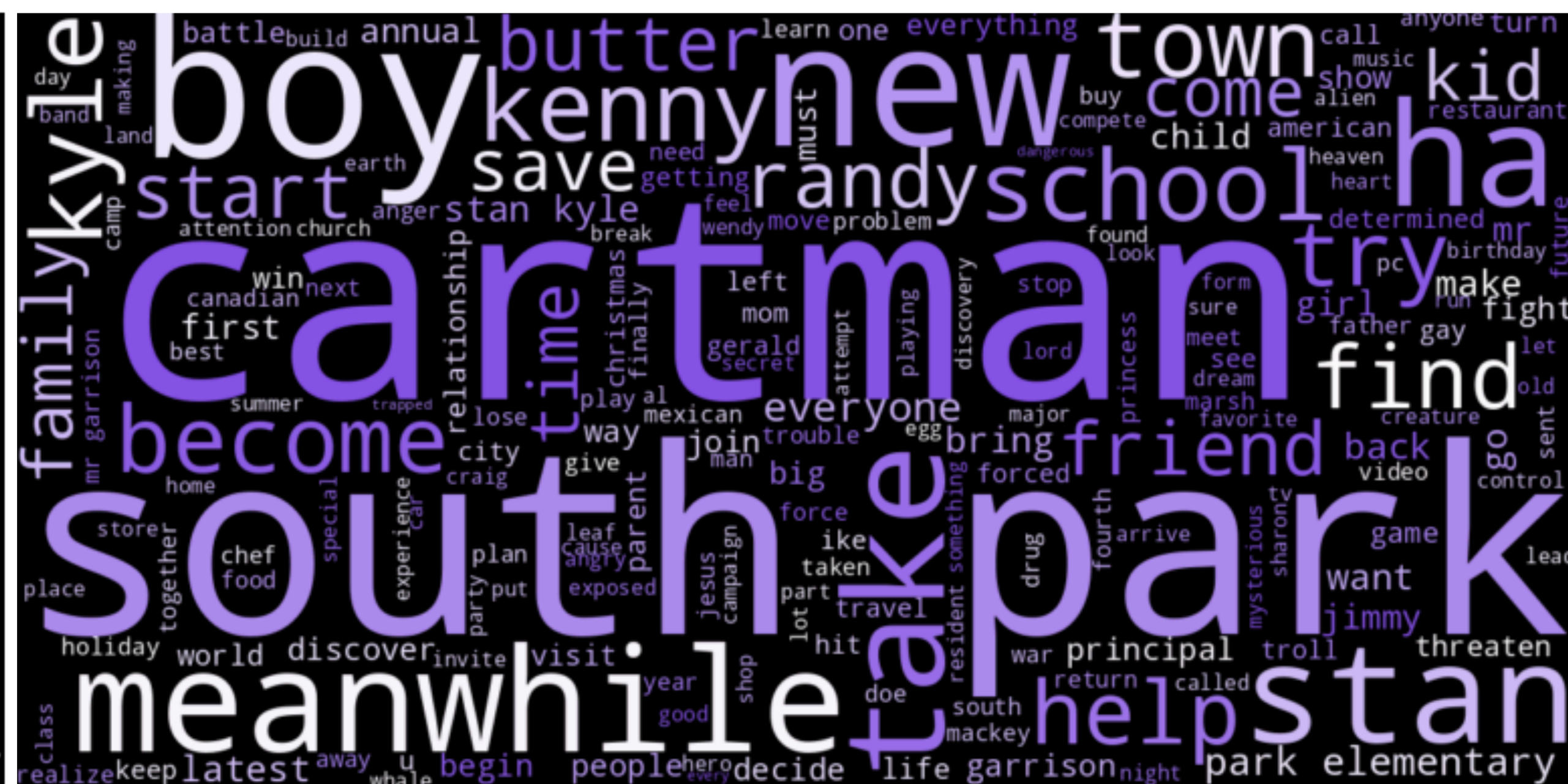
DATA EXPLORATION

Dialogue Vocabulary



- Character names are very prevalent

Summary Vocabulary



- **Character names still prevalent**
- **Narrative vocabulary**

METHODS



- **Extractive Summarization**
 - Gathers most important sentences from a text
 - Summary will contain actual dialogue
- **Abstractive Summarization**
 - Understands the semantics and structure of text
 - Summary will be paraphrased

EVALUATION: ROUGE SCORE

ROUGE compares a **produced summary** against a **reference summary** and generates a value between:

Bad  0.0 - 1.0  Great

MODELING & RESULTS

- TextRank Extractive Summarization: **Average ROUGE Score 0.09**
- Sequence-2-Sequence Abstractive Summarization: **Average ROUGE Score 0.05**
- Facebook's BART Abstractive Summarization: **Average ROUGE Score 0.11**



BART

Original Summary

The boys go to **China** to compete in a **dodgeball** championship. The town holds an appreciation week for the school nurse.

BART Summary

The South Park Cows Elementary School **Dodgeball** Team is going to the national finals. The Cows are on their way to the world championship in **China**.



RECOMMENDATIONS



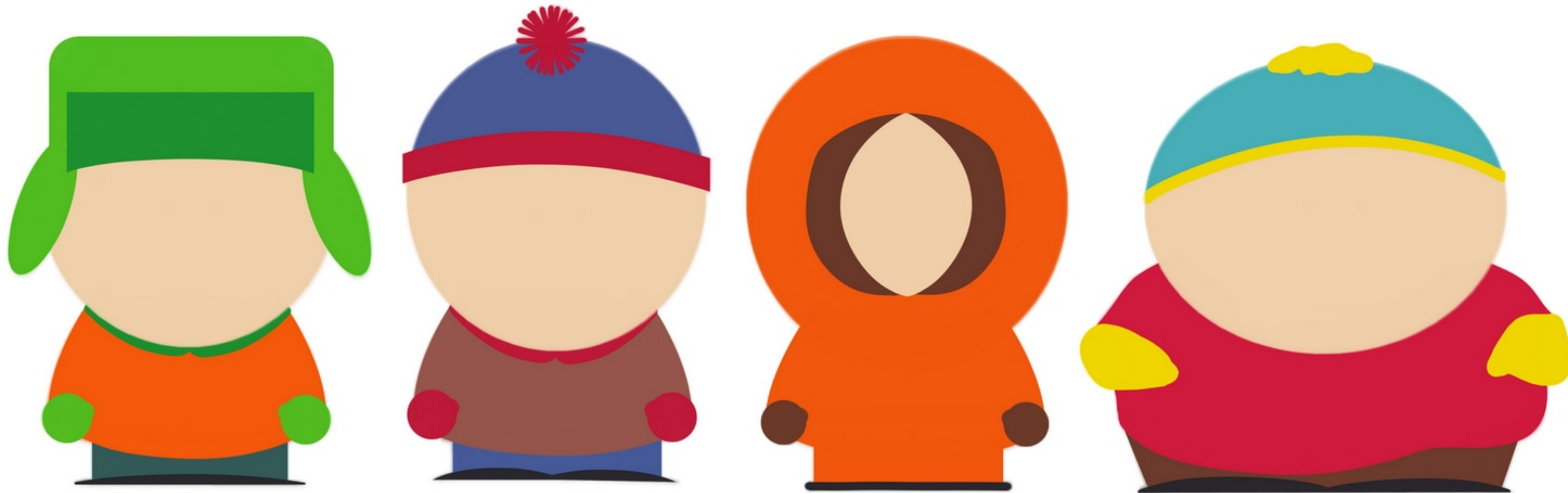
- Use pre-trained models to save you computing costs, time, and resources.
- Incorporate more dialogue from other texts and TV shows.



LIMITATIONS & FUTURE ANALYSIS

- **Include dialogue from other TV shows**
- **Explore different model architecture**
- **Fine-tune pre-trained models and save for future use**

THANK YOU



Questions?

You can reach me at: timothyjrabbitt@gmail.com

For more detailed analysis and methods behind this presentation, please visit: <https://github.com/trabbitt90/Text-Summarizer-of-South-Park-Episodes>