

Analyzing Podcast Episodes

Jeanette Pyne

Illinois Reboot Python 2.0

April 2022

Overview

- Goal of the project
- Process
- Results
- Future Goals

Goal of the Project



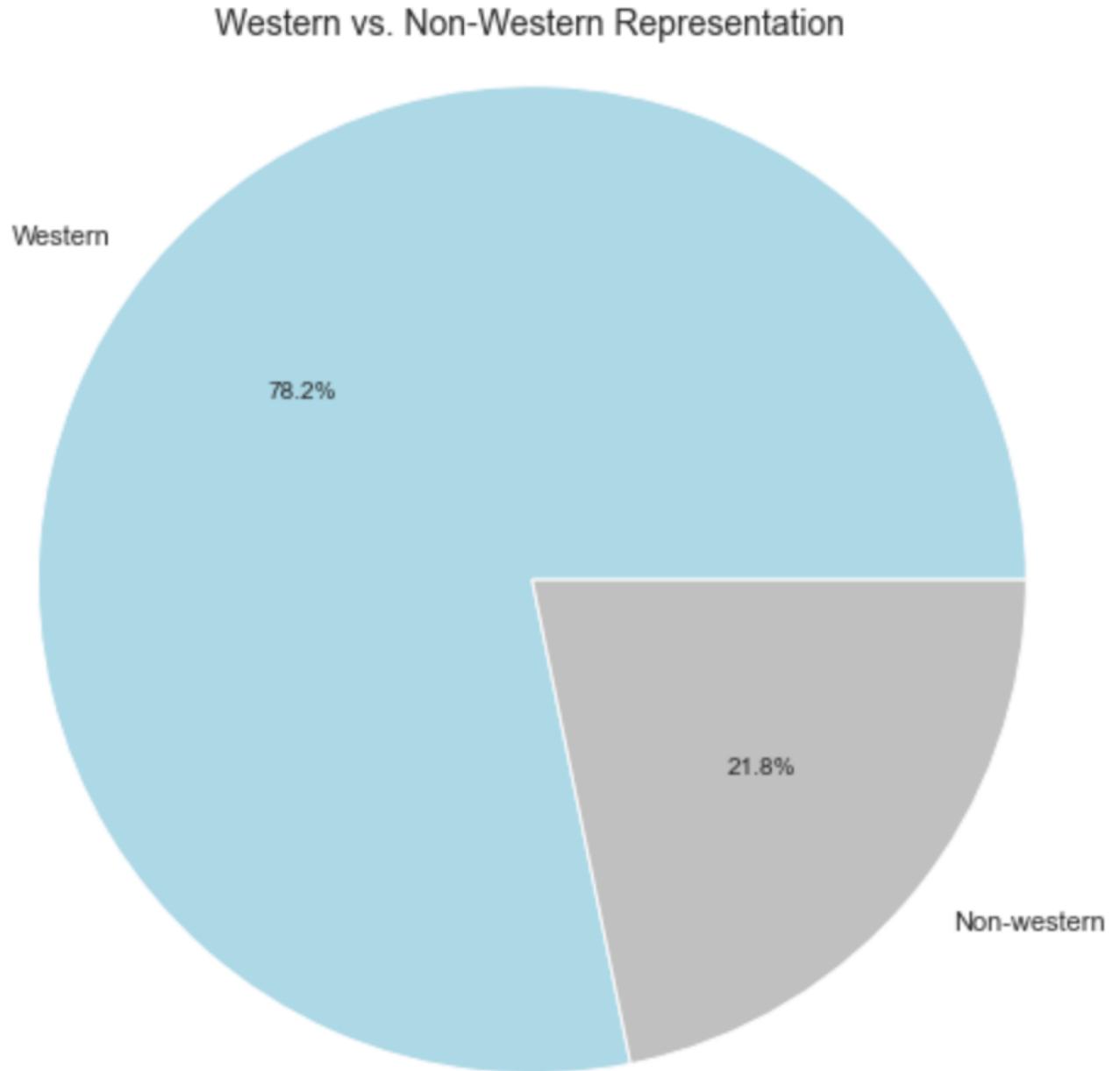
- Analyze episode descriptions from this history podcast
- How many of the episodes focus on...
 - Western vs. Non-western history?
 - Female vs Male history?
 - White people, minorities/people of color, and LGBTQ?
- Send findings to podcast hosts to make recommendations for future episodes

Process

- Tools
 - Python
 - web scraping, data preparation, data visualization
 - spaCy
 - Natural Language Processing (NLP) software
 - Named Entity Recognition (NER)
 - Excel
 - Hand coding countries into Western or Non-western countries**
 - Tableau
 - Data visualization

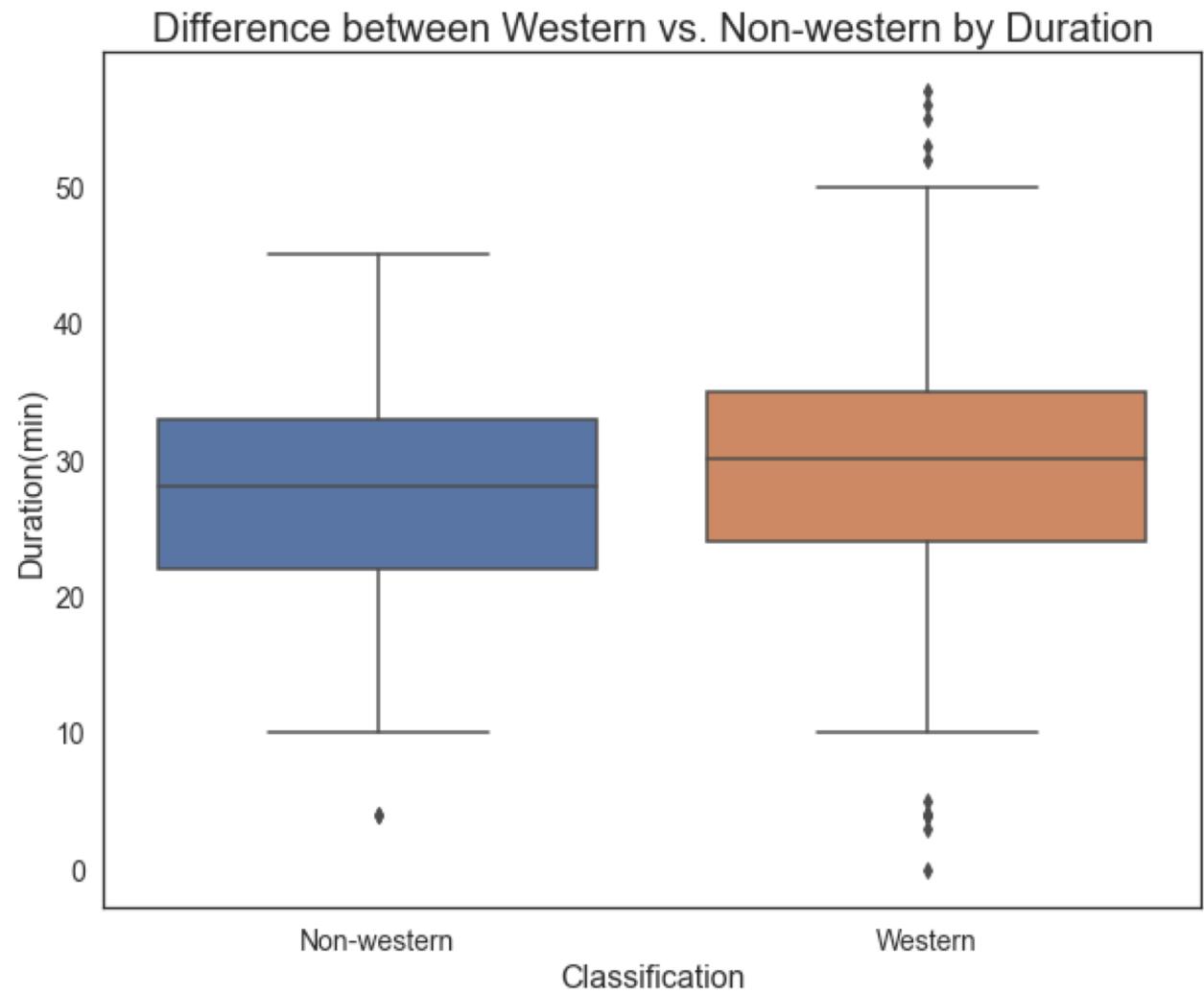
Results – Western vs. Non-Western Representation

- Focused on Western vs. Non-western representation
- Total episodes = 642

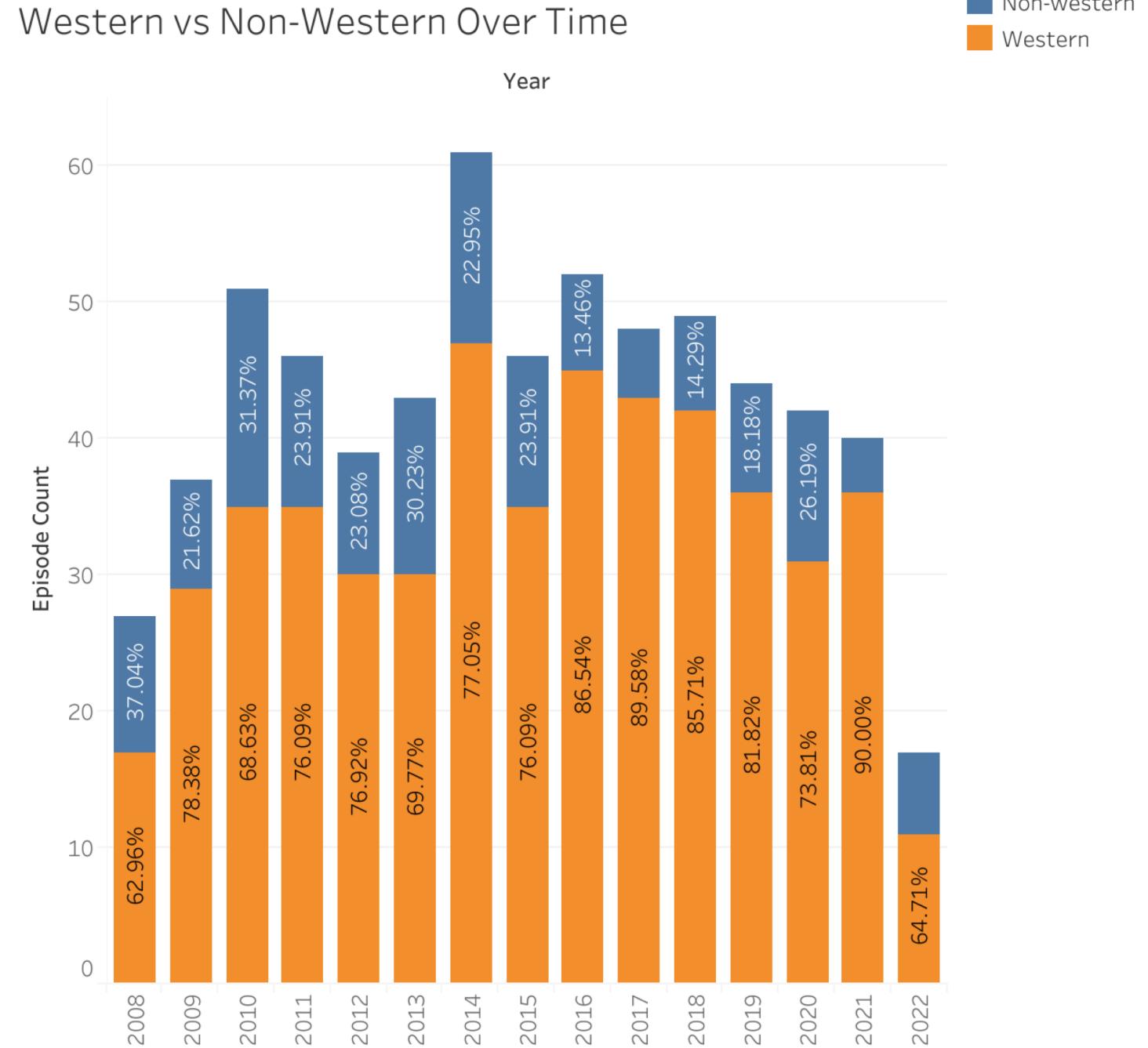


Results-- Box Plot and t-test

- My hypothesis: Difference in duration between Western and Non-western episodes
- Null hypothesis: There is no difference in duration
- Perform a t-test
 - P-value < 0.05
 - Difference is present and significant



Results—Stacked Bar Chart (Tableau)



Future Goals

- Western vs. Non-western
 - Network graph for relationship between Western and Non-western countries
- Other areas to analyze
 - Female/Male
 - Era/Period
 - White people vs. minorities (people of color and LGBTQ)
- Send report to the podcast hosts and offer recommendations

Thank you!

- Contact Information
 - Jeanette Pyne
 - jeanette.pyne@gmail.com