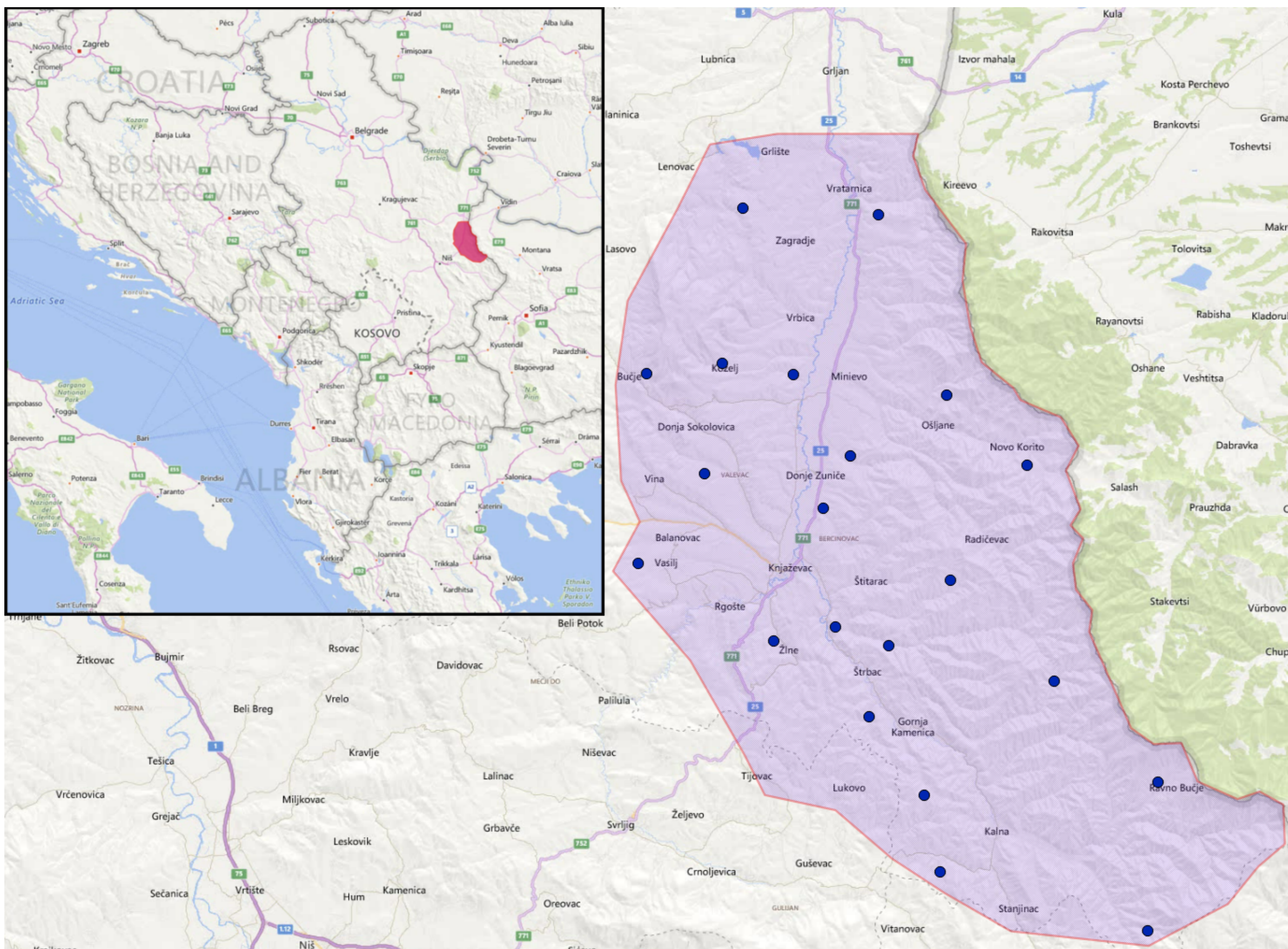# Definite Article Frequencies in Timok Dialect of Torlak

Teodora Vukovic

Institute for Slavic Linguistics,
CorpusLab, University of Zurich

Timok area in South-Eastern Serbia

Introduction

Determine how frequency of the definite article changes in contact with Bulgarian and Serbian. This feature is a part of standard Bulgarian, but not of standard Serbian.

> **Example**:
> Unuk=at   sadi   višnje=te.
> gradnson=DET plants cherries=DET
>
> 'The grandson is planting the cherries.'

Hypothesis

Definite article will be more frequent closer to the Bulgarian border and less frequent as the distance from the border increases (due to the influence of Bulgarian dialects and standard Serbian).

Sample and Method

Data processing

- **ExmarAlda**: 21 tanscripts of audio recordings
- **ReLDi Tagger**: the transcripts were annotated using Serbian part of speech and morphosyntactic tagger. Tags for nouns, adjectives and pronouns containing a definite article were manually corrected.
- **Pythons script**: total number of nouns and total number of articles were counted using a python script. Frequency of definite article was normalised in relation to the number of nouns.

Visualisation

- **Qgis**: normalized frequency counts were visualized using heatmaps in Qgis.

Hypothesis testing

- **Spearman Correlation Rank**:
  - Normalized frequency (FREQ) and Distance from Bulgarian Border (DIST)
  - Normalized frequency (FREQ) and longitude (LONG)
  - Normalized frequency (FREQ) and latitude (LAT)
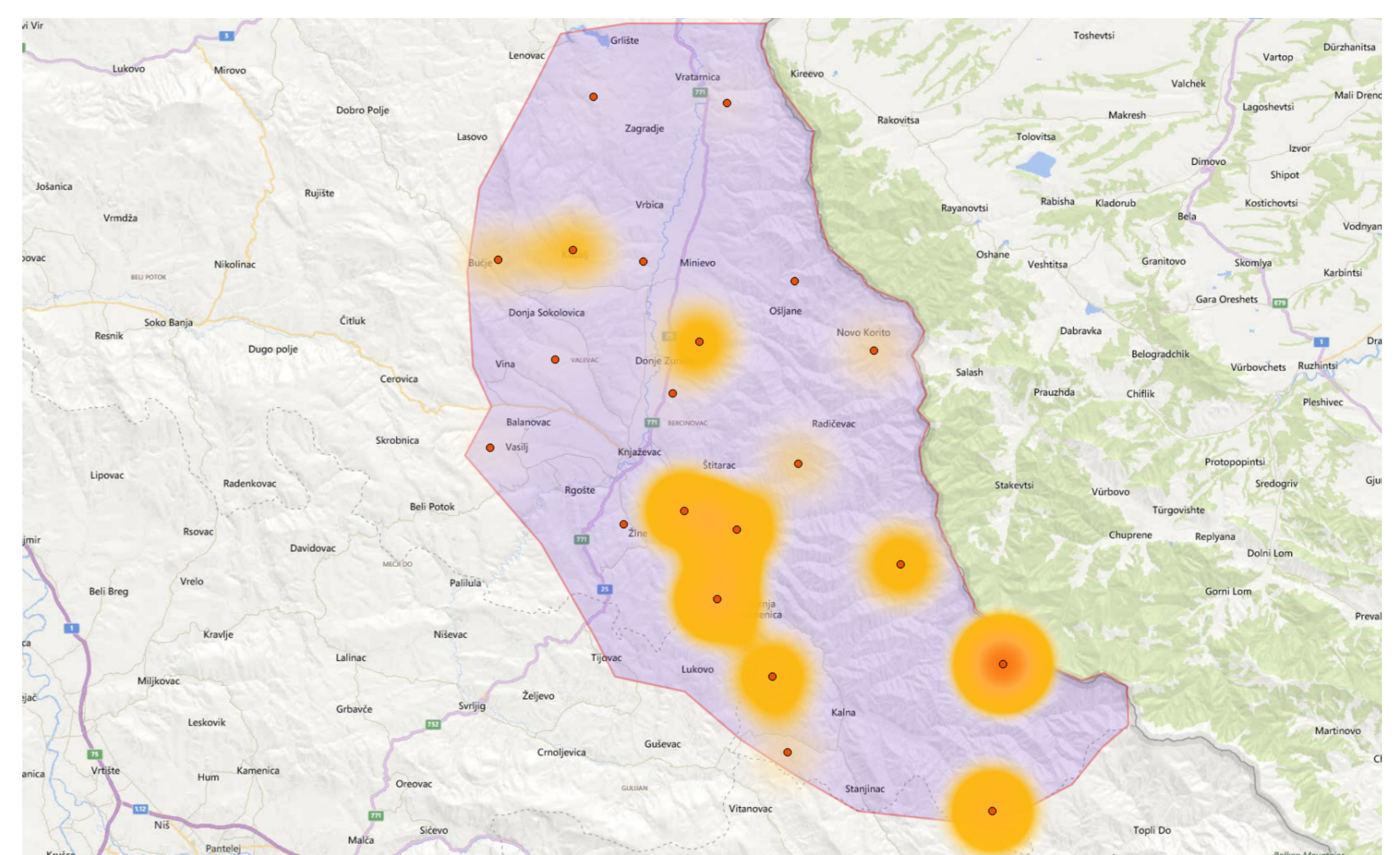
> **Example of annotated text:**
> ne   ne   Qz
> rabOti   rabotiti   Vm1rs
> nIgde   nigde   Rgp
> ženIcata ženica   Ncfsnyt
>
> 'That woman does not work anywhere.'

Results

Non-significant Spearman Correlation Rank results

- **FREQ and DIST** (N = 21)

  p-value = 0.5611

  rho = -0.1345029
- **FREQ and LONG** (N = 21)

  p-value = 0.03308

  rho = 0.4663852
- **FREQ and LAT** (N = 21)

  p-value = 0.03253

  rho = -0.4676843

Conclusion and Future Work

- **No significant score of Spearman Correlation Rank** between frequency and distance from Bulgarian border.
- **Emerging pattern against the hypothesis**: longitude seems to be playing a role rather than the distance from the border.
- **Future work**: more data points, other potential geographic factors, e.g. landscape, isolation.