Interactive Techniques for Visualising Categorical Data in Linguistics

David Trye, PhD Candidate in Computer Science

Supervised by Mark Apperley, David Bainbridge & Andreea Calude

KO TE TANGATA

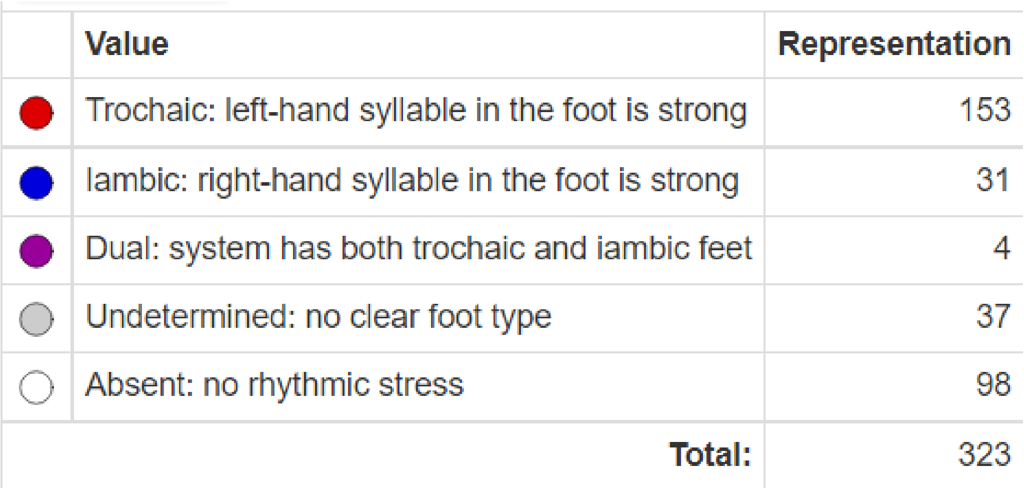
THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Presentation Aims

1. To introduce a novel visualisation tool called **Staircase Plots**
2. To encourage you to use this tool in your own analyses

Motivation

- Categorical data are **prevalent** in linguistics
  - The most common type of data in corpus linguistics (Stefanowitsch, 2020: 177)
  - Phonological, lexical, grammatical features (among others!)
- 192 WALS features (wals.info) with 2-28 categories
  - **Rhythm Type** (17A) has 5 categories, 323 items (languages)



	Value	Representation
🔴	Trochaic: left-hand syllable in the foot is strong	153
🔵	Iambic: right-hand syllable in the foot is strong	31
🟣	Dual: system has both trochaic and iambic feet	4
●	Undetermined: no clear foot type	37
○	Absent: no rhythmic stress	98
	Total:	323

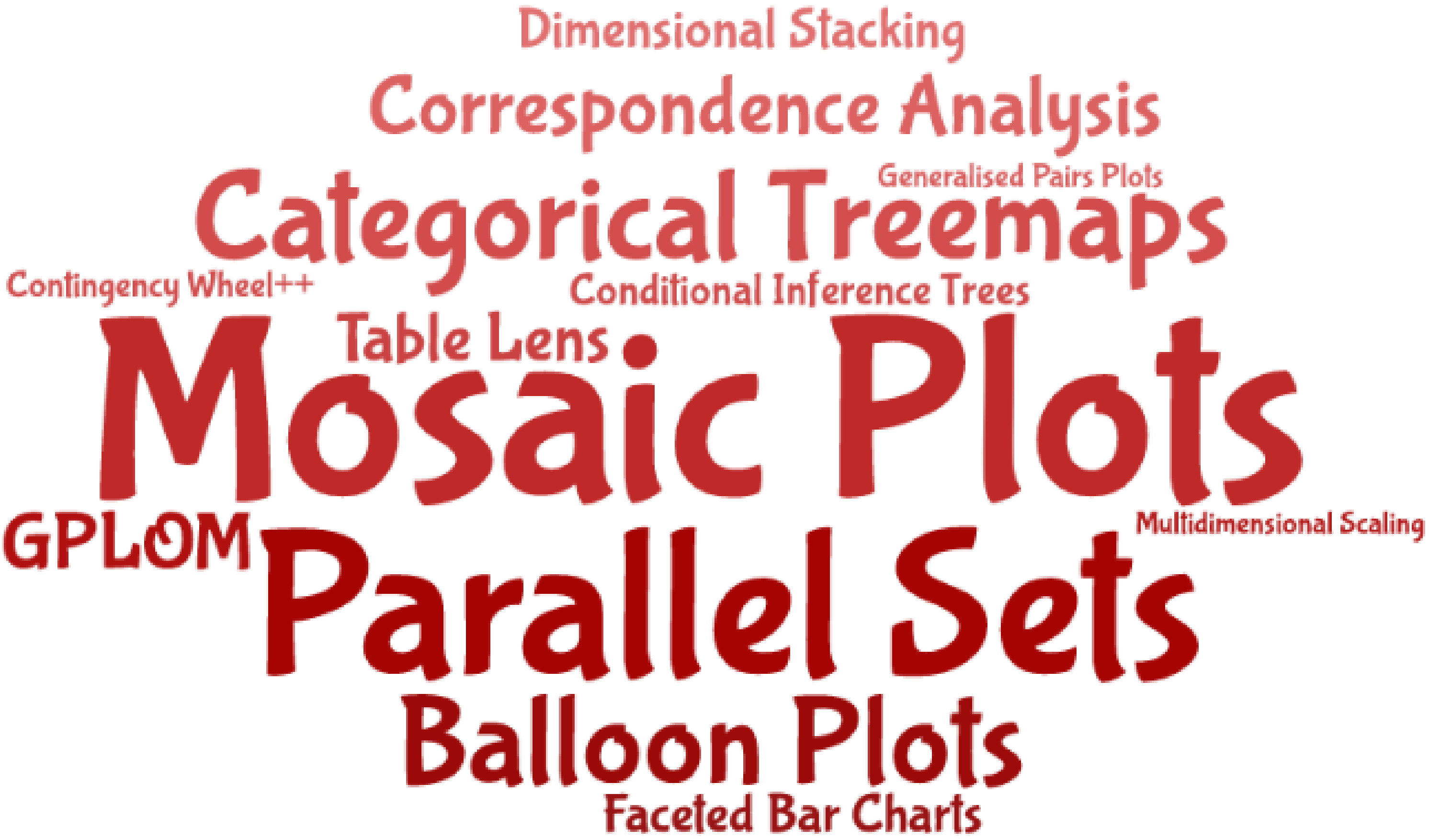
Motivation

- **Visualisation** can enhance linguistic analysis
  - Sanity checks
  - Anomaly detection
  - Knowledge discovery
  - Hypothesis testing
  - Statistical modelling
  - Presentation of results

Insights that might otherwise be missed!

- Few visualisation techniques effectively support **3+ categorical variables**
  - Limited scalability and interaction
  - Lack of user-friendly (no-code) tools available

Existing Techniques



Dimensional Stacking

Correspondence Analysis

Generalised Pairs Plots

Categorical Treemaps

Contingency Wheel++

Conditional Inference Trees

Table Lens

Mosaic Plots

GPLOM

Multidimensional Scaling

Parallel Sets

Balloon Plots

Faceted Bar Charts

Disclaimer

- Staircase Plots are currently **under development**
  - Design aspects are subject to change
  - Not available until next year

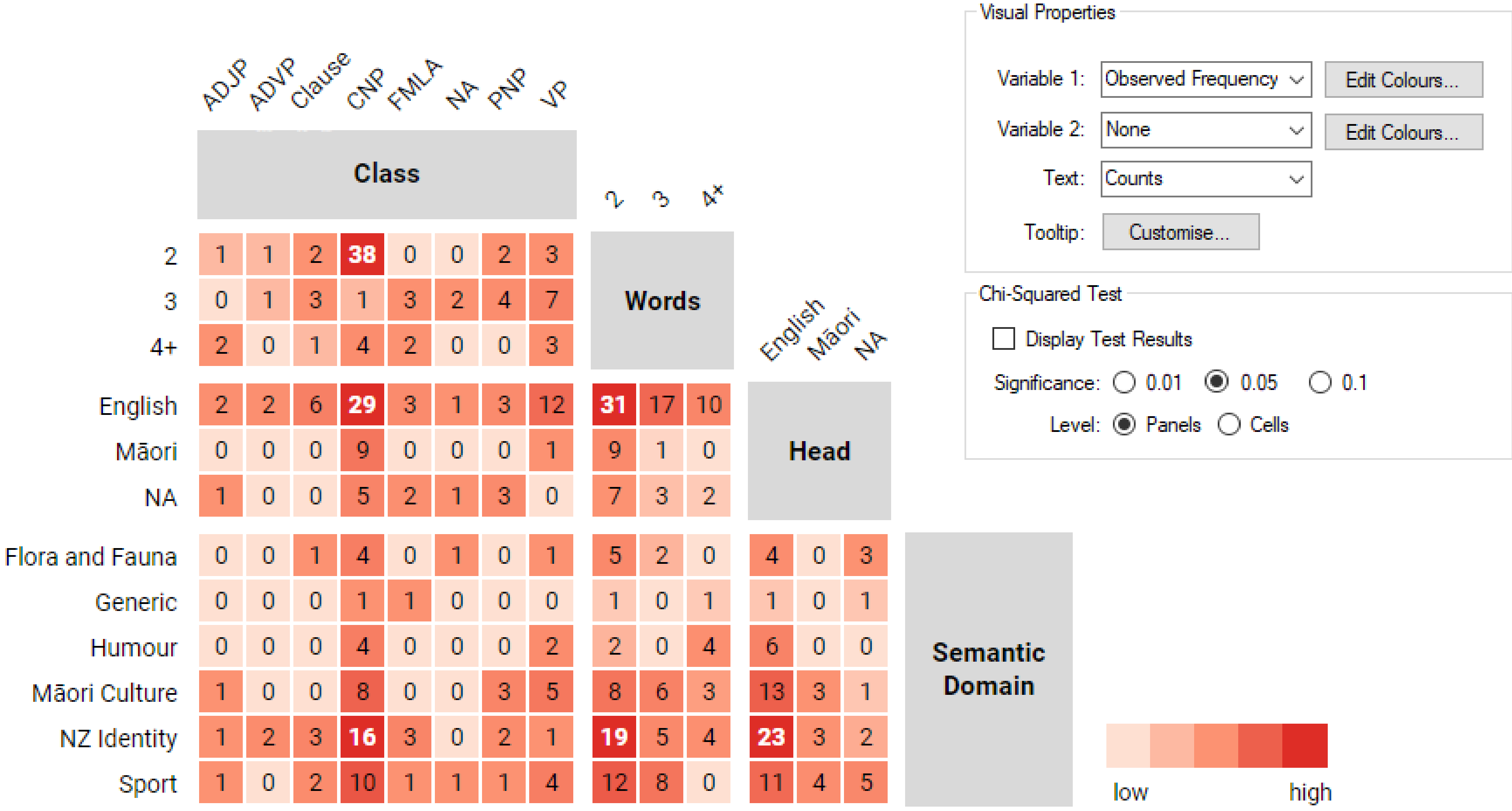
Dataset 1: Hybrid Hashtags

- 80 hashtags (rows) x 4 categorical variables (columns)
- Small **sample size**

Hashtag	Words (3)	Class (8)	Semantic Domain (6)	Head (3)
*#proud**kiwi***	2	CNP	Sport	Māori
#AotearoaNZ	3	PNP	NZ Identity	NA
⋮				
*#maori**pride***	2	CNP	Māori Culture	English

Source: Trye et al. (2020)

The Basics



		Class								Words			Head		
		ADJP	ADVP	Clause	CNP	FMLA	NA	PNP	VP	2	3	4+	English	Māori	NA
Words	2	1	1	2	38	0	0	2	3						
	3	0	1	3	1	3	2	4	7						
	4+	2	0	1	4	2	0	0	3						
Head	English	2	2	6	29	3	1	3	12	31	17	10			
	Māori	0	0	0	9	0	0	0	1	9	1	0			
	NA	1	0	0	5	2	1	3	0	7	3	2			
Semantic Domain	Flora and Fauna	0	0	1	4	0	1	0	1	5	2	0	4	0	3
	Generic	0	0	0	1	1	0	0	0	1	0	1	1	0	1
	Humour	0	0	0	4	0	0	0	2	2	0	4	6	0	0
	Māori Culture	1	0	0	8	0	0	3	5	8	6	3	13	3	1
	NZ Identity	1	2	3	16	3	0	2	1	19	5	4	23	3	2
	Sport	1	0	2	10	1	1	1	4	12	8	0	11	4	5

low high

Visual Properties

Variable 1: Observed Frequency

Variable 2: None

Text: Counts

Tooltip: Customise...

Chi-Squared Test

☐ Display Test Results

Significance: ○ 0.01 ● 0.05 ○ 0.1

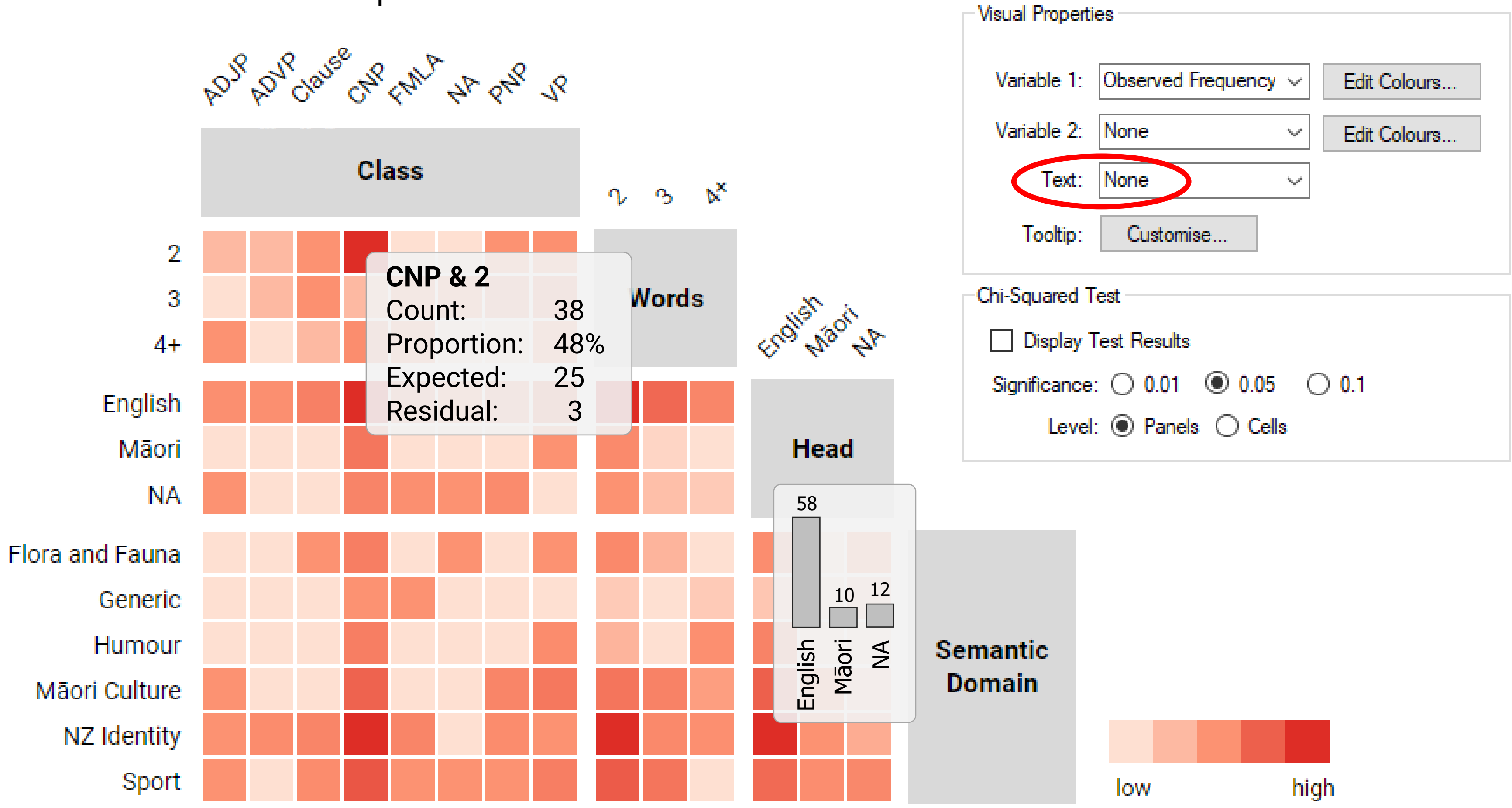
Level: ● Panels ○ Cells

Why this approach?

- “A good **starting point** for any data exploration is a simple summary table” (Brezina, 2018: 108)
  - An even better starting point is a heatmap!
- “It is always useful to do **cross-tabulation** of all categorical predictors and the response before beginning your analysis in order to detect configurations with zero frequencies or a large number of cells with **very low frequencies**” (Levshina, 2015: 273)

Removing Text Labels

- Easier to perceive general patterns
- Interactive tooltips reveal details on demand



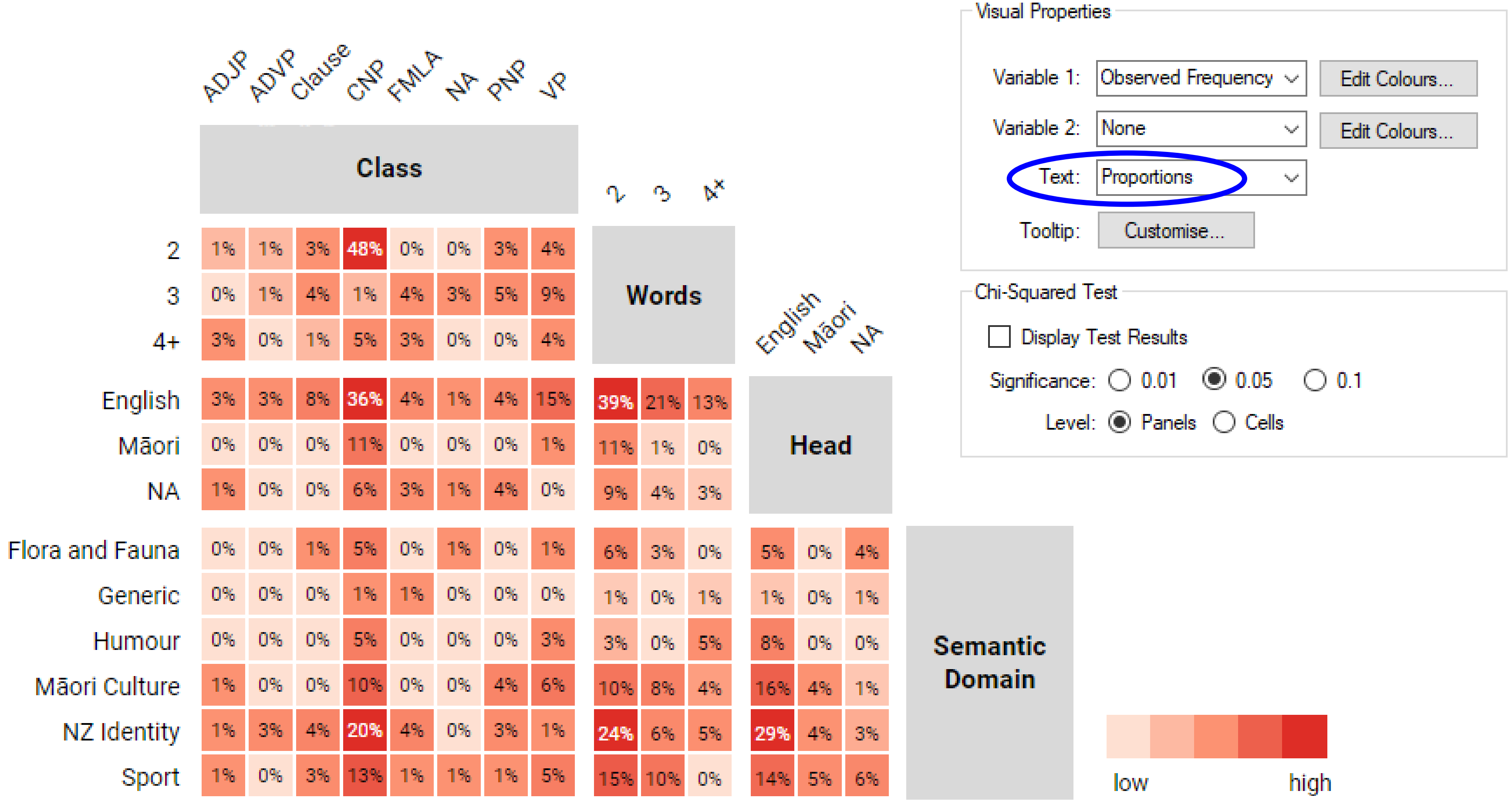
CNP & 2
Count: 38
Proportion: 48%
Expected: 25
Residual: 3

Visual Properties
Variable 1: Observed Frequency
Variable 2: None
Text: None
Tooltip: Customise...

Chi-Squared Test
□ Display Test Results
Significance: ○ 0.01 ● 0.05 ○ 0.1
Level: ● Panels ○ Cells

Proportions

- Cells show joint probability, P(X∩Y), where X and Y are the categories on each axis



		Class								Words			Head		
		ADJP	ADVP	Clause	CNP	FMLA	NA	PNP	VP	2	3	4+	English	Māori	NA
Words	2	1%	1%	3%	48%	0%	0%	3%	4%						
	3	0%	1%	4%	1%	4%	3%	5%	9%						
	4+	3%	0%	1%	5%	3%	0%	0%	4%						
Head	English	3%	3%	8%	36%	4%	1%	4%	15%	39%	21%	13%			
	Māori	0%	0%	0%	11%	0%	0%	0%	1%	11%	1%	0%			
	NA	1%	0%	0%	6%	3%	1%	4%	0%	9%	4%	3%			
Semantic Domain	Flora and Fauna	0%	0%	1%	5%	0%	1%	0%	1%	6%	3%	0%	5%	0%	4%
	Generic	0%	0%	0%	1%	1%	0%	0%	0%	1%	0%	1%	1%	0%	1%
	Humour	0%	0%	0%	5%	0%	0%	0%	3%	3%	0%	5%	8%	0%	0%
	Māori Culture	1%	0%	0%	10%	0%	0%	4%	6%	10%	8%	4%	16%	4%	1%
	NZ Identity	1%	3%	4%	20%	4%	0%	3%	1%	24%	6%	5%	29%	4%	3%
	Sport	1%	0%	3%	13%	1%	1%	1%	5%	15%	10%	0%	14%	5%	6%

low high

Visual Properties

Variable 1: Observed Frequency

Variable 2: None

Text: Proportions

Tooltip:

Chi-Squared Test

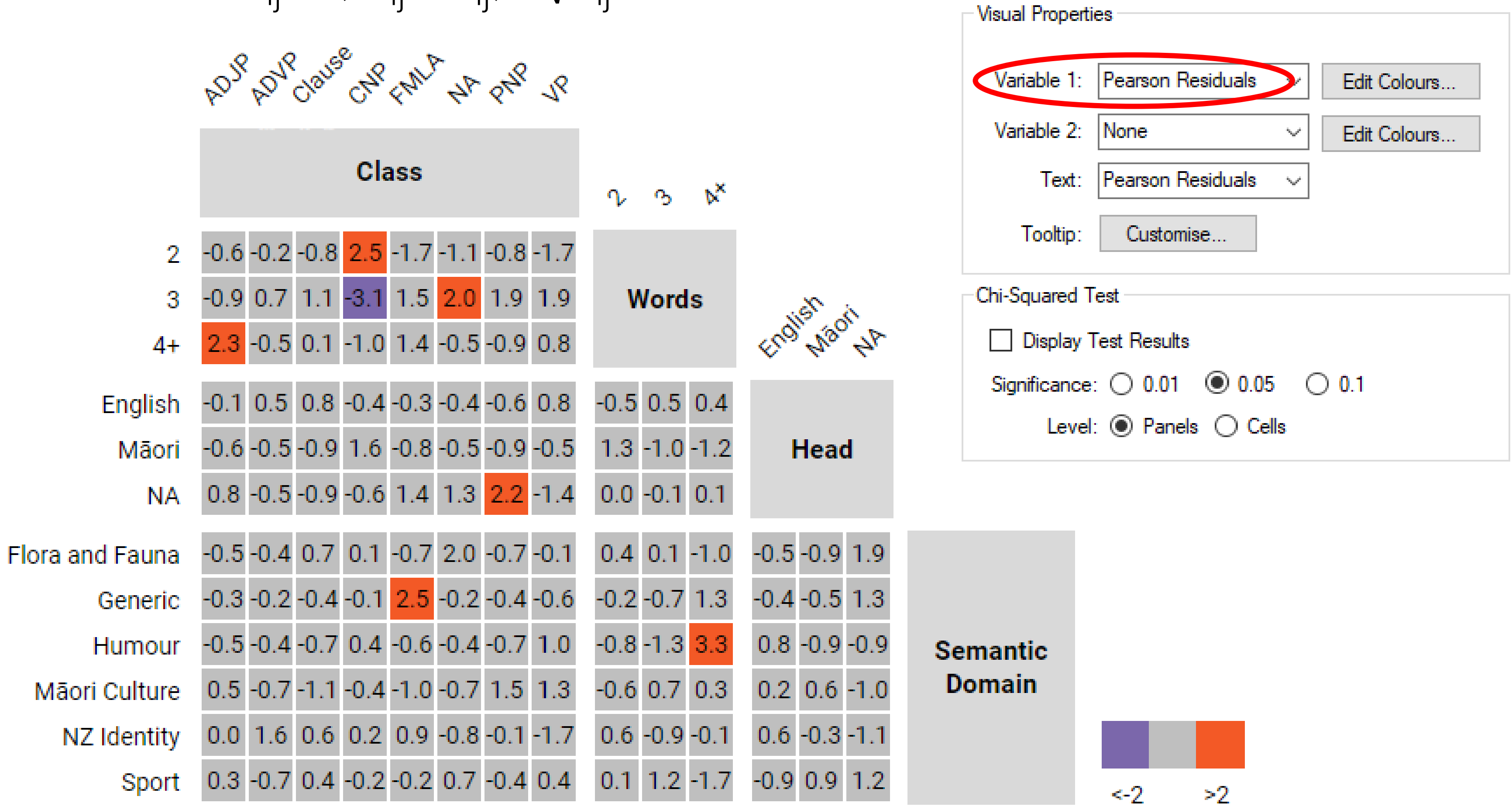
☐ Display Test Results

Significance: ○ 0.01 ● 0.05 ○ 0.1

Level: ● Panels ○ Cells

Pearson Residuals

- Non-grey cells (>|2|) correspond to sig. residuals when $\alpha \approx 0.05$
- Formula: $r_{ij} = (O_{ij} - E_{ij}) / \sqrt{E_{ij}}$



		Class								Words			Head		
		ADJP	ADVP	Clause	CNP	FMLA	NA	PNP	VP	2	3	4+	English	Māori	NA
Words	2	-0.6	-0.2	-0.8	2.5	-1.7	-1.1	-0.8	-1.7						
	3	-0.9	0.7	1.1	-3.1	1.5	2.0	1.9	1.9						
	4+	2.3	-0.5	0.1	-1.0	1.4	-0.5	-0.9	0.8						
Head	English	-0.1	0.5	0.8	-0.4	-0.3	-0.4	-0.6	0.8	-0.5	0.5	0.4			
	Māori	-0.6	-0.5	-0.9	1.6	-0.8	-0.5	-0.9	-0.5	1.3	-1.0	-1.2			
	NA	0.8	-0.5	-0.9	-0.6	1.4	1.3	2.2	-1.4	0.0	-0.1	0.1			
Semantic Domain	Flora and Fauna	-0.5	-0.4	0.7	0.1	-0.7	2.0	-0.7	-0.1	0.4	0.1	-1.0	-0.5	-0.9	1.9
	Generic	-0.3	-0.2	-0.4	-0.1	2.5	-0.2	-0.4	-0.6	-0.2	-0.7	1.3	-0.4	-0.5	1.3
	Humour	-0.5	-0.4	-0.7	0.4	-0.6	-0.4	-0.7	1.0	-0.8	-1.3	3.3	0.8	-0.9	-0.9
	Māori Culture	0.5	-0.7	-1.1	-0.4	-1.0	-0.7	1.5	1.3	-0.6	0.7	0.3	0.2	0.6	-1.0
	NZ Identity	0.0	1.6	0.6	0.2	0.9	-0.8	-0.1	-1.7	0.6	-0.9	-0.1	0.6	-0.3	-1.1
	Sport	0.3	-0.7	0.4	-0.2	-0.2	0.7	-0.4	0.4	0.1	1.2	-1.7	-0.9	0.9	1.2

🟪 <-2 🟧 >2

Visual Properties

Variable 1: Pearson Residuals [Edit Colours...]

Variable 2: None [Edit Colours...]

Text: Pearson Residuals

Tooltip: [Customise...]

Chi-Squared Test

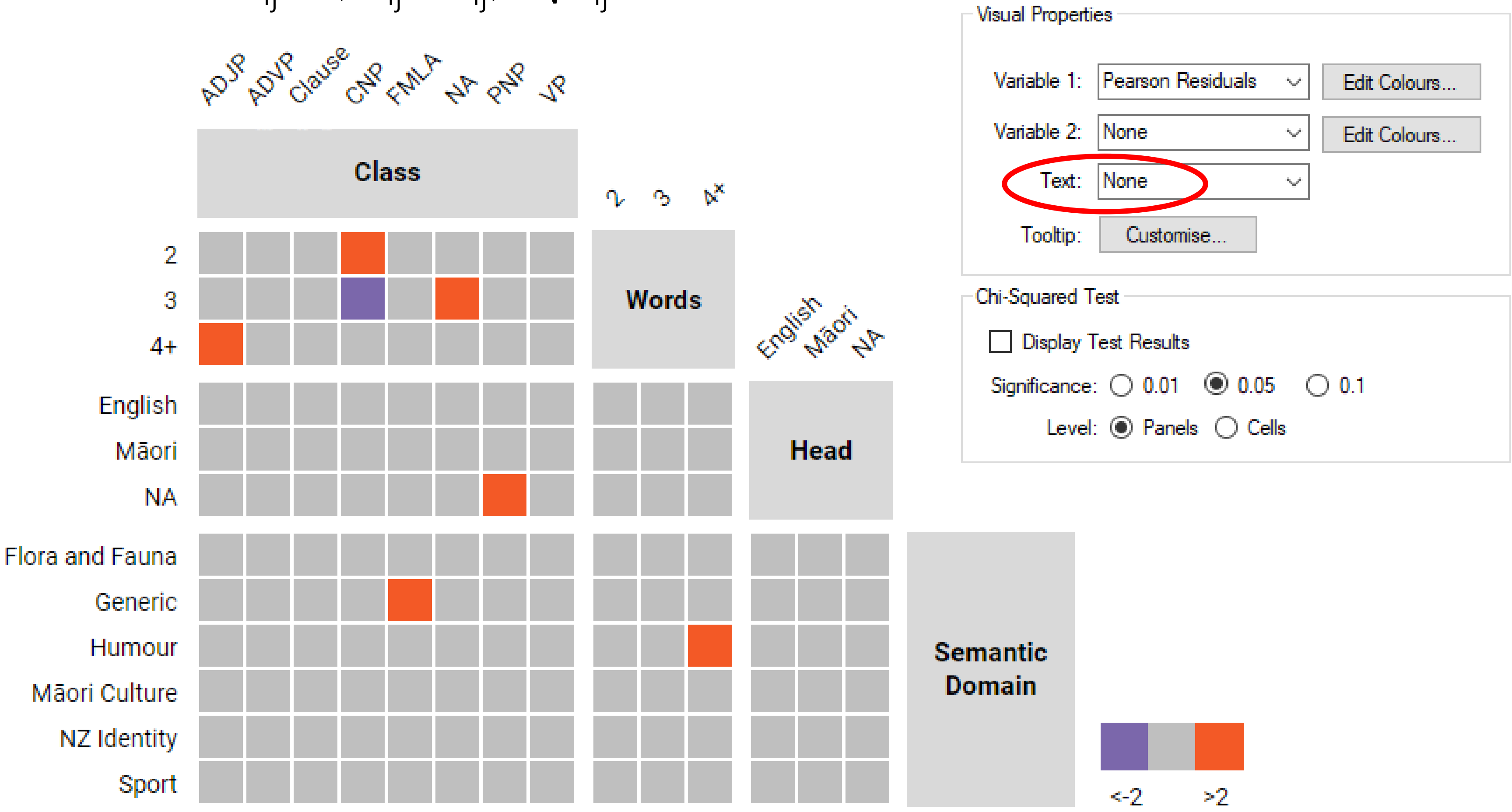
□ Display Test Results

Significance: ○ 0.01 ● 0.05 ○ 0.1

Level: ● Panels ○ Cells

Pearson Residuals

- Non-grey cells (>|2|) correspond to sig. residuals when α ≈ 0.05
- Formula: $r_{ij} = (O_{ij} - E_{ij}) / \sqrt{E_{ij}}$



	ADJP	ADVP	Clause	CNP	FMLA	NA	PNP	VP
2	▒	▒	▒	🟧	▒	▒	▒	▒
3	▒	▒	▒	🟪	▒	🟧	▒	▒
4+	🟧	▒	▒	▒	▒	▒	▒	▒
English	▒	▒	▒	▒	▒	▒	▒	▒
Māori	▒	▒	▒	▒	▒	▒	▒	▒
NA	▒	▒	▒	▒	▒	▒	🟧	▒
Flora and Fauna	▒	▒	▒	▒	▒	▒	▒	▒
Generic	▒	▒	▒	▒	🟧	▒	▒	▒
Humour	▒	▒	▒	▒	▒	▒	▒	▒
Māori Culture	▒	▒	▒	▒	▒	▒	▒	▒
NZ Identity	▒	▒	▒	▒	▒	▒	▒	▒
Sport	▒	▒	▒	▒	▒	▒	▒	▒

	2	3	4+
English	▒	▒	▒
Māori	▒	▒	▒
NA	▒	▒	▒
Flora and Fauna	▒	▒	▒
Generic	▒	▒	▒
Humour	▒	▒	🟧
Māori Culture	▒	▒	▒
NZ Identity	▒	▒	▒
Sport	▒	▒	▒

	English	Māori	NA
Flora and Fauna	▒	▒	▒
Generic	▒	▒	▒
Humour	▒	▒	▒
Māori Culture	▒	▒	▒
NZ Identity	▒	▒	▒
Sport	▒	▒	▒

Visual Properties

Variable 1: Pearson Residuals ⌄ Edit Colours...

Variable 2: None ⌄ Edit Colours...

Text: None ⌄

Tooltip: Customise...

Chi-Squared Test

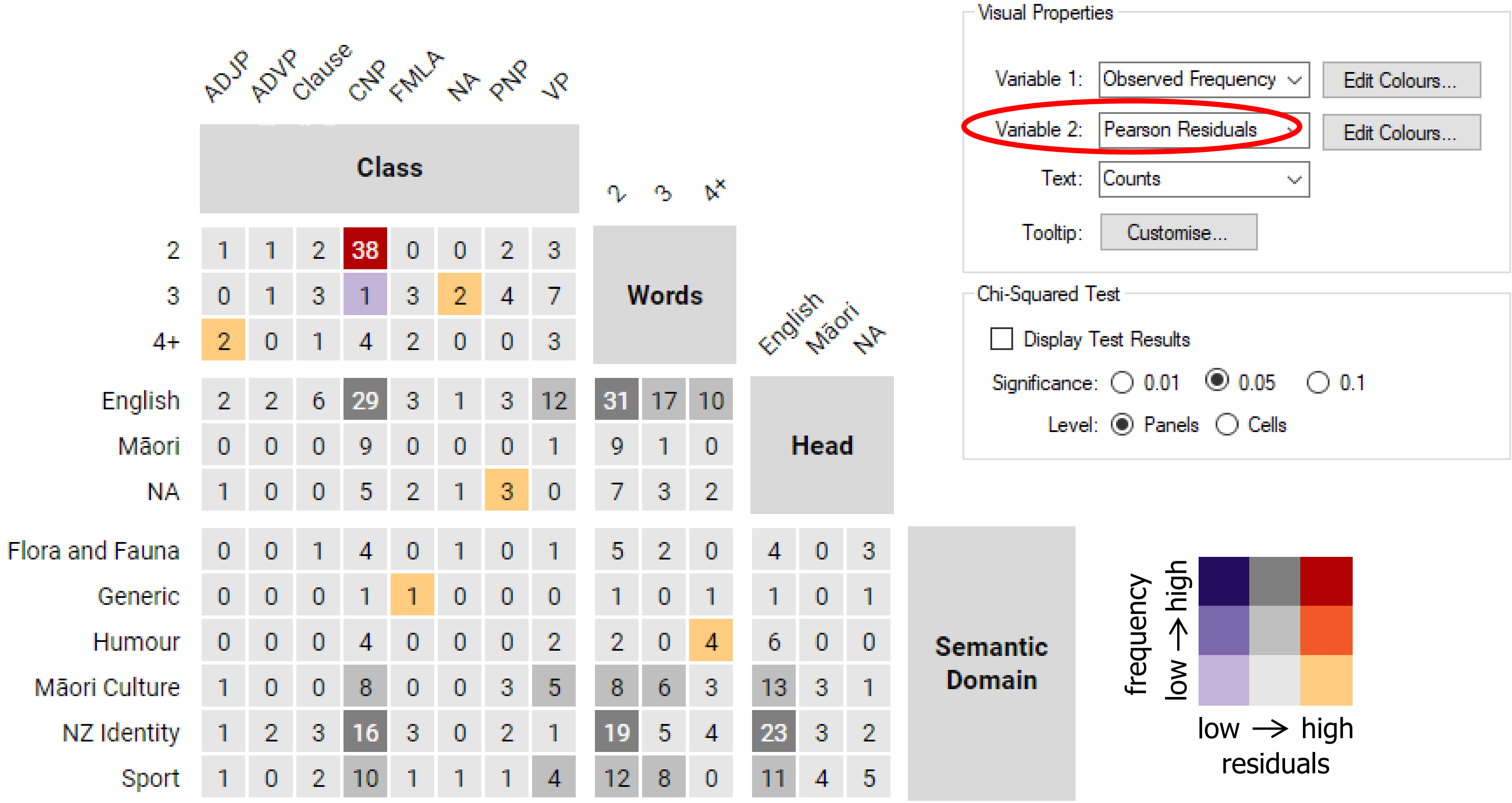
☐ Display Test Results

Significance: ○ 0.01 ● 0.05 ○ 0.1

Level: ● Panels ○ Cells

Bivariate Colour Scheme

- Show both counts (intensity) *and* residuals (colour)



		Class								Words			Head		
		ADJP	ADVP	Clause	CNP	FMLA	NA	PNP	VP	2	3	4+	English	Māori	NA
Words	2	1	1	2	38	0	0	2	3						
	3	0	1	3	1	3	2	4	7						
	4+	2	0	1	4	2	0	0	3						
Head	English	2	2	6	29	3	1	3	12	31	17	10			
	Māori	0	0	0	9	0	0	0	1	9	1	0			
	NA	1	0	0	5	2	1	3	0	7	3	2			
Semantic Domain	Flora and Fauna	0	0	1	4	0	1	0	1	5	2	0	4	0	3
	Generic	0	0	0	1	1	0	0	0	1	0	1	1	0	1
	Humour	0	0	0	4	0	0	0	2	2	0	4	6	0	0
	Māori Culture	1	0	0	8	0	0	3	5	8	6	3	13	3	1
	NZ Identity	1	2	3	16	3	0	2	1	19	5	4	23	3	2
	Sport	1	0	2	10	1	1	1	4	12	8	0	11	4	5

Visual Properties

Variable 1: Observed Frequency
Variable 2: Pearson Residuals
Text: Counts

Chi-Squared Test

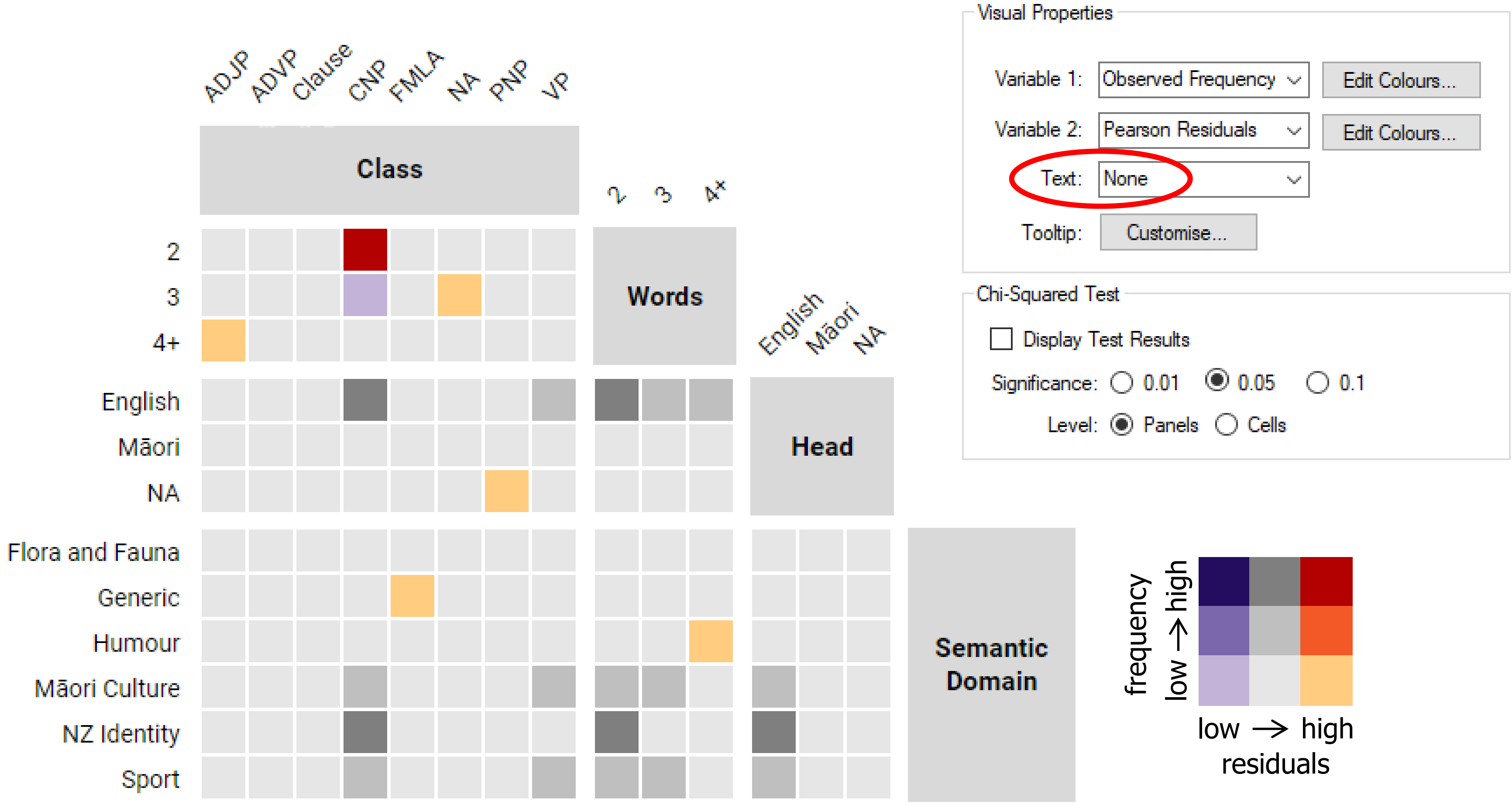
☐ Display Test Results

Significance: ○ 0.01 ● 0.05 ○ 0.1

Level: ● Panels ○ Cells

Bivariate Colour Scheme

- Show both counts (intensity) *and* residuals (colour)



Visual Properties

Variable 1: Observed Frequency — Edit Colours...

Variable 2: Pearson Residuals — Edit Colours...

Text: None

Tooltip: Customise...

Chi-Squared Test

☐ Display Test Results

Significance: ○ 0.01 ● 0.05 ○ 0.1

Level: ● Panels ○ Cells

Chi-Squared Test

- Staircase Plots provide built-in support for the chi-squared test of independence
  - Used to determine whether there is an **association** between two categorical variables
- Ability to calculate & display results for all pairs of variables that satisfy the basic **test conditions**
  - Panels coloured according to strength of association
    - **Effect size** measured using Cramer’s V
- Advantages:
  - Removes burden of manual computation
  - Visually reinforces correct interpretation
  - All results conveniently displayed in one place

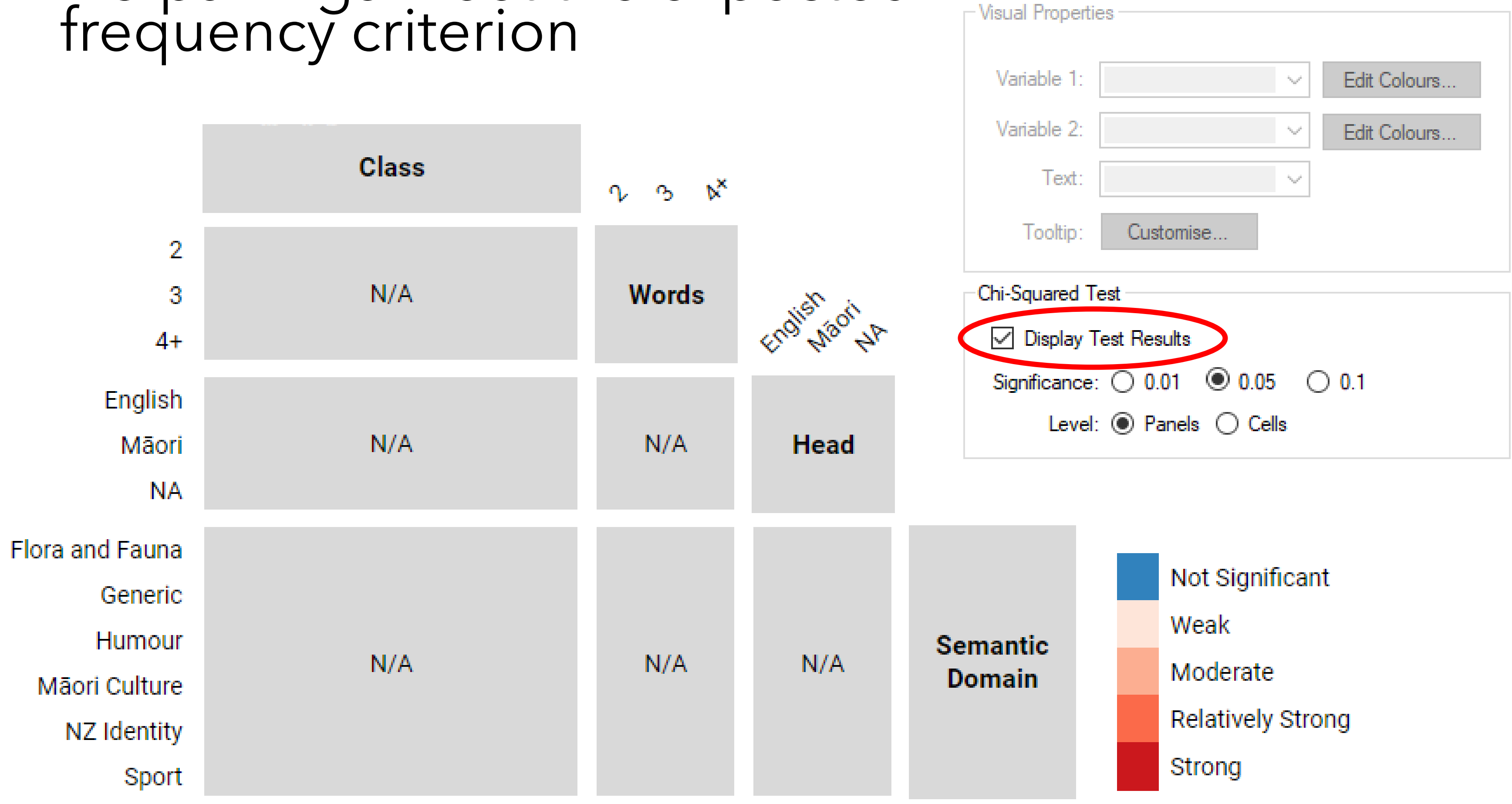
Chi-Squared Test Conditions

1. Nominal (preferred) or ordinal variables
    - Quantitative variables can be binned
2. Independent observations
    - Requires manual verification
3. Mutually-exclusive categories
    - Each observation contributes to one cell per panel
4. Expected frequency >1 in *all* cells and >5 in at least 80% of cells
    - Requires decent sample size
    - Typically at least 5x number of cells

	A	B	
C	X	-	Row
D	-	-	-
	Col	-	*N*

Chi-Squared Test

- **Insufficient sample size** for this dataset!
- No pairings meet the expected frequency criterion



Class				
N/A	Words			
N/A	N/A	Head		
N/A	N/A	N/A	Semantic Domain	

Visual Properties

Variable 1: Edit Colours...

Variable 2: Edit Colours...

Text:

Tooltip: Customise...

Chi-Squared Test

☑ Display Test Results

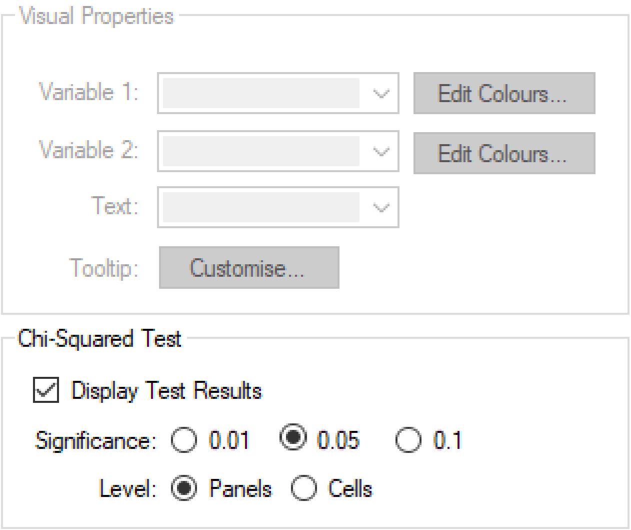
Significance: ○ 0.01 ● 0.05 ○ 0.1

Level: ● Panels ○ Cells

Chi-Squared Test

- Example of a larger dataset (N= 2,201)
- Each panel reports the test statistic, (degrees of freedom), p-value & Cramer's V

	crew 3rd 1st 2nd			
	Class	adult child		
adult child	5.59 (3) $p < 0.001$, $\varphi_c = 0.139$	**Age**	male female	
male female	349.91 (3) $p < 0.001$, $\varphi_c = 0.399$	27.12 (1) $p < 0.001$ $\varphi_c = 0.111$	**Sex**	
no yes	190.4 (3) $p < 0.001$ $\varphi_c = 0.294$	20.96 (1) $p < 0.001$ $\varphi_c = 0.098$	456.87 (1) $p < 0.001$ $\varphi_c = 0.456$	**Survived**


Visual Properties

Variable 1: Edit Colours...

Variable 2: Edit Colours...

Text:

Tooltip: Customise...

Chi-Squared Test

☑ Display Test Results

Significance: ○ 0.01 ● 0.05 ○ 0.1

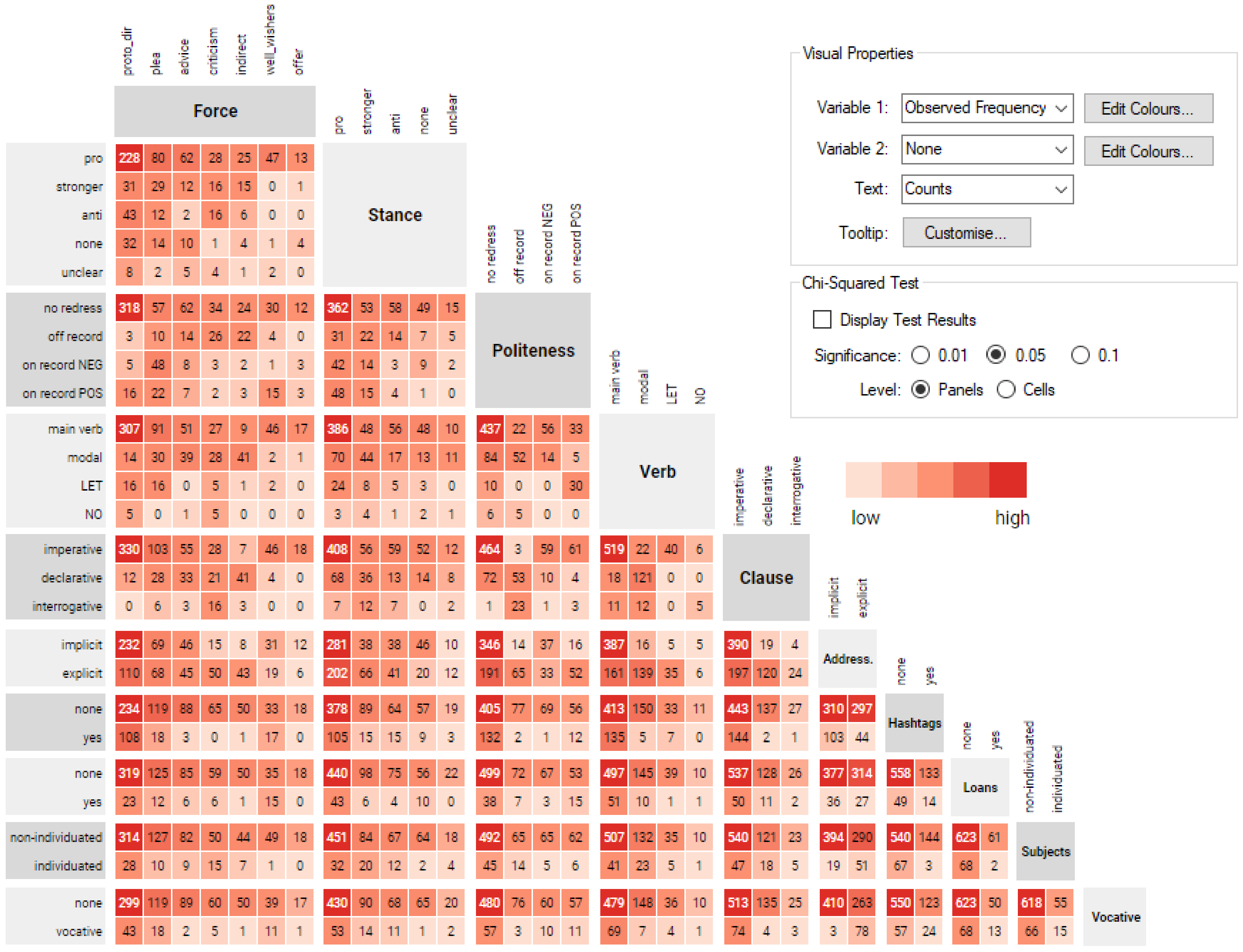
Level: ● Panels ○ Cells

- 🟦 Not Significant
- 🟥 Weak
- 🟥 Moderate
- 🟥 Relatively Strong
- 🟥 Strong

Dataset 2: Covid Directives (Burnette & Calude, 2022)

- 754 directives (rows) from tweets containing *#covid19nz*
- 10 variables (columns)

Variable	Categories
Stance (5)	against, pro, for stronger measures, neutral, unclear
Force (7)	advice, criticism, indirect, offer, plea, prototypical, well wishers
Politeness (4)	no redress, on record negative, on record positive, off record
Verb (4)	let, main verb, modal, no
Clause (3)	declarative, imperative, interrogative
Addressees (2)	explicit, implicit
Hashtags (2)	none, yes
Loanwords (2)	none, yes
Subjects (2)	individuated, non-individuated
Vocative (2)	none, yes



		Force							Stance					Politeness				Verb				Clause			Address.		Hashtags		Loans		Subjects	
		proto_dir	plea	advice	criticism	indirect	well_wishers	offer	pro	stronger	anti	none	unclear	no redress	off record	on record NEG	on record POS	main verb	modal	LET	NO	imperative	declarative	interrogative	implicit	explicit	none	yes	none	yes	non-individuated	individuated
Stance	pro	228	80	62	28	25	47	13																								
	stronger	31	29	12	16	15	0	1																								
	anti	43	12	2	16	6	0	0																								
	none	32	14	10	1	4	1	4																								
	unclear	8	2	5	4	1	2	0																								
Politeness	no redress	318	57	62	34	24	30	12	362	53	58	49	15																			
	off record	3	10	14	26	22	4	0	31	22	14	7	5																			
	on record NEG	5	48	8	3	2	1	3	42	14	3	9	2																			
	on record POS	16	22	7	2	3	15	3	48	15	4	1	0																			
Verb	main verb	307	91	51	27	9	46	17	386	48	56	48	10	437	22	56	33															
	modal	14	30	39	28	41	2	1	70	44	17	13	11	84	52	14	5															
	LET	16	16	0	5	1	2	0	24	8	5	3	0	10	0	0	30															
	NO	5	0	1	5	0	0	0	3	4	1	2	1	6	5	0	0															
Clause	imperative	330	103	55	28	7	46	18	408	56	59	52	12	464	3	59	61	519	22	40	6											
	declarative	12	28	33	21	41	4	0	68	36	13	14	8	72	53	10	4	18	121	0	0											
	interrogative	0	6	3	16	3	0	0	7	12	7	0	2	1	23	1	3	11	12	0	5											
Address.	implicit	232	69	46	15	8	31	12	281	38	38	46	10	346	14	37	16	387	16	5	5	390	19	4								
	explicit	110	68	45	50	43	19	6	202	66	41	20	12	191	65	33	52	161	139	35	6	197	120	24								
Hashtags	none	234	119	88	65	50	33	18	378	89	64	57	19	405	77	69	56	413	150	33	11	443	137	27	310	297						
	yes	108	18	3	0	1	17	0	105	15	15	9	3	132	2	1	12	135	5	7	0	144	2	1	103	44						
Loans	none	319	125	85	59	50	35	18	440	98	75	56	22	499	72	67	53	497	145	39	10	537	128	26	377	314	558	133				
	yes	23	12	6	6	1	15	0	43	6	4	10	0	38	7	3	15	51	10	1	1	50	11	2	36	27	49	14				
Subjects	non-individuated	314	127	82	50	44	49	18	451	84	67	64	18	492	65	65	62	507	132	35	10	540	121	23	394	290	540	144	623	61		
	individuated	28	10	9	15	7	1	0	32	20	12	2	4	45	14	5	6	41	23	5	1	47	18	5	19	51	67	3	68	2		
Vocative	none	299	119	89	60	50	39	17	430	90	68	65	20	480	76	60	57	479	148	36	10	513	135	25	410	263	550	123	623	50	618	55
	vocative	43	18	2	5	1	11	1	53	14	11	1	2	57	3	10	11	69	7	4	1	74	4	3	3	78	57	24	68	13	66	15

Visual Properties

Variable 1: Observed Frequency

Variable 2: None

Text: Counts

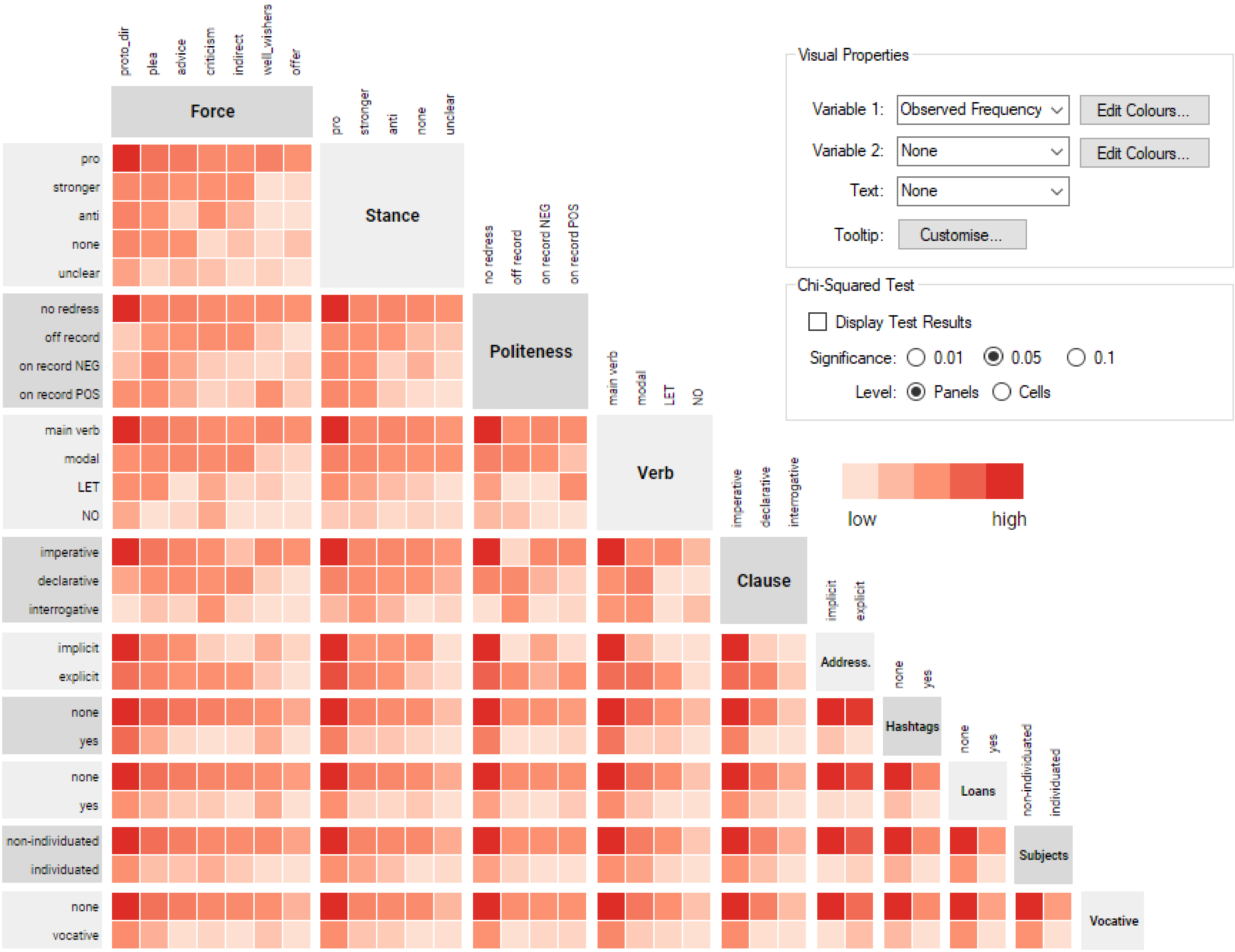
Tooltip:

Chi-Squared Test

☐ Display Test Results

Significance: ○ 0.01 ● 0.05 ○ 0.1

Level: ● Panels ○ Cells



Visual Properties

Variable 1: Observed Frequency — Edit Colours...

Variable 2: None — Edit Colours...

Text: None

Tooltip: Customise...

Chi-Squared Test

□ Display Test Results

Significance: ○ 0.01 ● 0.05 ○ 0.1

Level: ● Panels ○ Cells

Key Limitations

- Inner variables are **split** across columns and rows
  - Displaying only half the matrix saves space but makes comparison with other variables difficult
- Layout restricts total **number of categories** that can be displayed
  - Don't want multiple variables with 10+ categories
  - Exact limit varies according to screen resolution
- **Loss of precision** when using bivariate colour maps
  - Fewer distinct shades for each variable
- Not optimised for **ordinal data**
  - Chi-squared test doesn't consider ordering information

Interactive Features – Coming Soon!

- Display selected items in **scrollable table**
- **Associative highlighting** for categories (rows/columns) & variables (related panels)
  - Related: search feature
- Flexible **re-ordering** of categories & variables
  - Alphabetically, by frequency/cardinality, manually via drag-and-drop
- Basic **data transformations**
  - Collapse/expand existing categories
  - Add/remove variables
  - Filter by selection

Help needed!

- Online participants needed for a **user study** about Staircase Plots
  - ~1 hour in Feb/March 2023
- Please fill out this quick Google Form (name + email) if you might be interested in taking part
- Thank you!

https://forms.gle/evL7j3jed8ZgfgVM6

Ngā pātai?

Contact me

David Trye

dgt12@students.waikato.ac.nz

(Or talk to me on Stream 1 during the breaks!)

References (1)

- Benzécri, J. P. (1992). *Correspondence analysis handbook.* CRC Press LLC.
- Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide.* Cambridge University Press.
- Burnette, J., & Calude, A. S. (2022). Wake up New Zealand! Directives, politeness and stance in Twitter #Covid19NZ posts. *Journal of Pragmatics*, *196*, 6-23.
- Emerson, J. W., Green, W. A., Schloerke, B., Crowley, J., Cook, D., Hofmann, H., & Wickham, H. (2013). The generalized pairs plot. *Journal of Computational and Graphical Statistics*, *22*(1), 79-91.
- Friendly, M. (1994). Mosaic displays for multi-way contingency tables. Journal of the American Statistical Association, 89(425), 190-200.
- Friendly, M. (1999). Extending mosaic displays: Marginal, conditional, and partial views of categorical data. *Journal of Computational and graphical Statistics*, *8*(3), 373-395.
- Grinstein, G., Trutschl, M., & Cvek, U. (2001, August). High-dimensional visualizations. In *Proceedings of the Visual Data Mining Workshop*, KDD (Vol. 2, p. 120).
- Hartigan, J. A., & Kleiner, B. (1981). Mosaics for contingency tables. In *Computer science and statistics: Proceedings of the 13th symposium on the interface* (pp. 268-273). Springer, New York, NY.
- Im, J. F., McGuffin, M. J., & Leung, R. (2013). GPLOM: the generalized plot matrix for visualizing multidimensional multivariate data. *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2606-2614.
- Jain, N., & Warnes, G. R. (2006). Balloon plot. *The Newsletter of the R Project Volume 6/2*, May 2006, 6, 35.
- Kosara, R., Bendix, F., & Hauser, H. (2006). Parallel sets: Interactive exploration and visual analysis of categorical data. *IEEE transactions on visualization and computer graphics*, *12*(4), 558-568.
- Kolatchm, E., & Weinstein, B. (2001). CatTrees: Dynamic visualization of categorical data using treemaps. http://www.cs.umd.edu/class/spring2001/cmsc838b/project/kolatch_weinstein/index.html

References (2)

- Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. John Benjamins Publishing Company.

- Levshina, N. (2020). Conditional inference trees and random forests. In *A practical handbook of corpus linguistics* (pp. 611-643). Springer, Cham.

- Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., & Pfister, H. (2014). UpSet: visualization of intersecting sets. *IEEE transactions on visualization and computer graphics*, *20*(12), 1983-1992.

- Mead, A. (1992). Review of the development of multidimensional scaling methods. Journal of the Royal Statistical Society: Series D (The Statistician), 41(1), 27-39.

- Rao, R., & Card, S. K. (1994, April). The table lens: merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 318-322).

- Rocha, M. M. N., & da Silva, C. G. (2018). Heatmap matrix: A multidimensional data visualization technique. In *Proceedings of the 31st Conference on Graphics, Patterns and Images (SIBGRAPI)*.

- Rocha, M. M. N., & da Silva, C. G. (2022). Heatmap matrix: Using reordering, discretization and filtering resources to assist multidimensional data analysis.

- Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*. Language Science Press.

- Theus, M. (2002). Interactive data visualization using Mondrian. *Journal of Statistical Software*, *7*, 1-9.

- Trye, D. (2022, April 11-14). Visualising multivariate categorical data. In *Proceedings of the IEEE Pacific Visualization Symposium (PacificVis)*, Tsukuba, Japan.

- Trye, D., Calude, A. S., Bravo-Marquez, F., & Keegan, T. T. (2020). Hybrid hashtags: #YouKnowYoureAKiwiWhen your tweet contains Māori and English. *Frontiers in artificial intelligence*, *3*, 15.

- Valdivia, P., Buono, P., Plaisant, C., Dufournaud, N., & Fekete, J. D. (2019). Analyzing dynamic hypergraphs with parallel aggregated ordered hypergraph visualization. *IEEE transactions on visualization and computer graphics*, *27*(1), 1-13.