Mikko Hakala
Tanja Säily
Turo Vartiainen

Variation in pronoun frequencies in early English letters

Gender-based or relationship-based?

Gender differences in pronoun use

- Women use more pronouns than men in LME and EModE letters (Säily et al. 2011)
  - Similar observations for PDE (Rayson et al. 1997; Argamon et al. 2003)
- Differences explained in terms of gendered styles
  - A high frequency of 1st and 2nd person pronouns indicates an involved style (Biber 1988, 1995; Biber & Burges 2000)
  - ➢ Women's style is more involved, whereas men's style is more informational

Relationship between sender and recipient?

- Säily et al. (2011) mostly ignored the influence of register, audience design, and corpus balance
- We explore the effect of audience design on pronoun use in 17th and 18th century letters

> To what extent are pronoun frequencies affected by the relationship between the sender and the recipient of the letter?

- Focus on 1st and 2nd person pronouns *I* and *you*, two markers of involved style

Material

Sample from the *Corpora of Early English Correspondence*

- Letters from 1600–1800
  - 3.85 million words (0.94 million by women, 2.91 million by men)
- Categories based on the sender-recipient relationship
  - Letters written by or addressed to
    1. Members of the nuclear family (FN)
    2. Other family members (FO)
    3. Family servants (FS)
    4. Close friends (TC)
    5. All other acquaintances (T)

Relationship categories in CEEC

Number of words in different relationship categories

Relationship Category	Women		Men		Total	
FN	445,678	47%	739,724	25%	1,185,402	31%
FO	164,303	17%	332,818	11%	497,121	13%
FS	3,486	0%	25,069	1%	28,555	1%
TC	179,378	19%	709,580	24%	888,958	23%
T	150,341	16%	1,103,555	38%	1,253,896	33%
Total	943,186		2,910,746		3,853,932	

Family letters studied

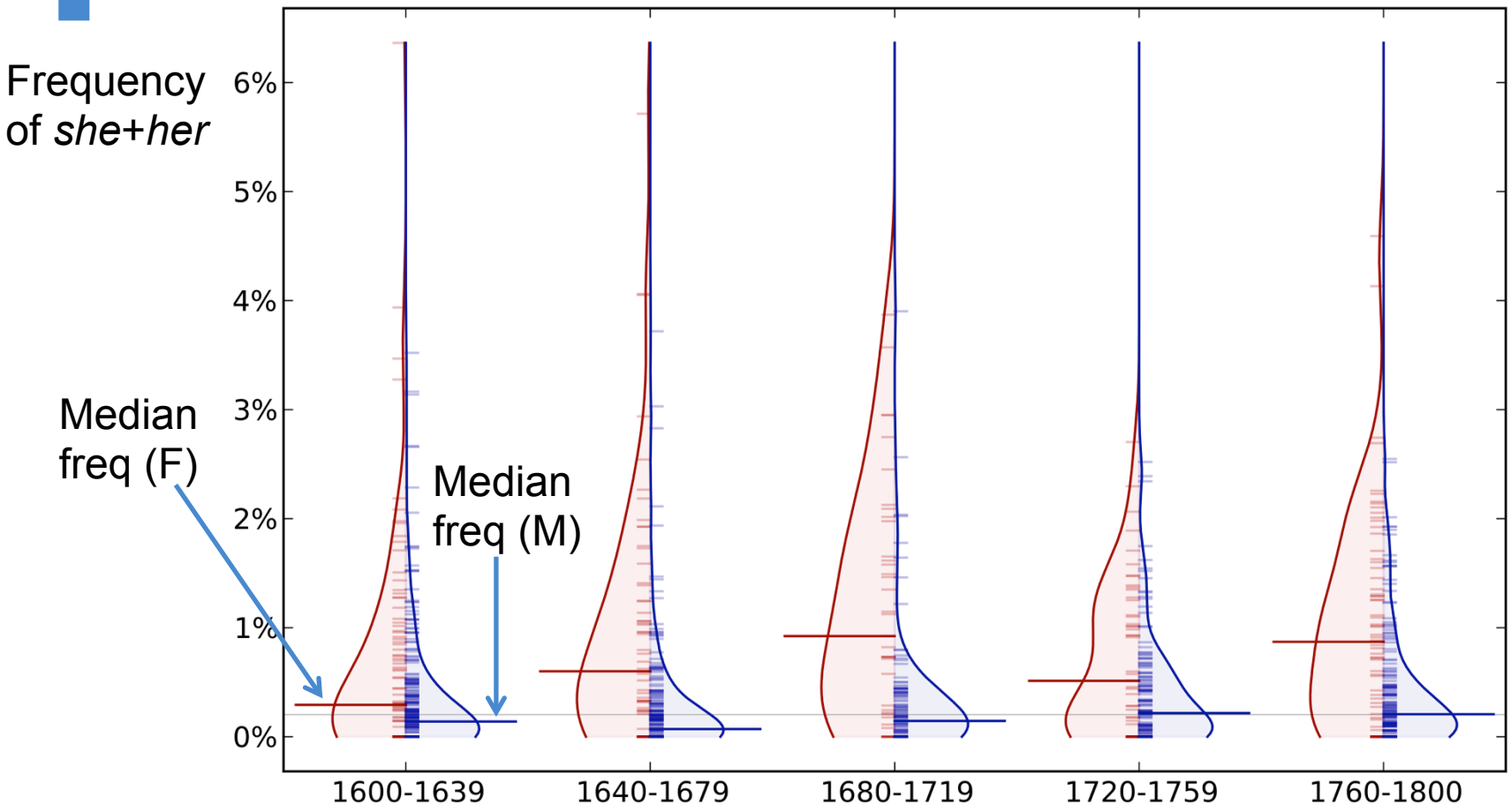
Number of words in the family letters data set

	1600–1679		1680–1800		Entire period	
husband > wife	101,015	40%	15,716	7%	116,731	24%
wife > husband	31,502	12%	51,910	22%	83,412	17%
father > son	26,756	10%	39,560	17%	66,316	13%
son > father	29,090	11%	94,931	40%	124,021	25%
mother > son	41,734	16%	18,902	8%	60,636	12%
son > mother	25,039	10%	15,381	7%	40,420	8%
Total	255,136		236,400		491,536	

Outliers and corpus balance

- Aggregate normalised frequencies (e.g. M, F) problematic due to outliers
  - ➢ Use **smaller samples**
  - Sample: person's letters from a 20-year period addressed to a specific relationship category
- Compare categories based on **median frequency** of samples
  - Visualisation: **beanplots** (Kampstra 2008) with median frequencies

Visualisation: beanplots

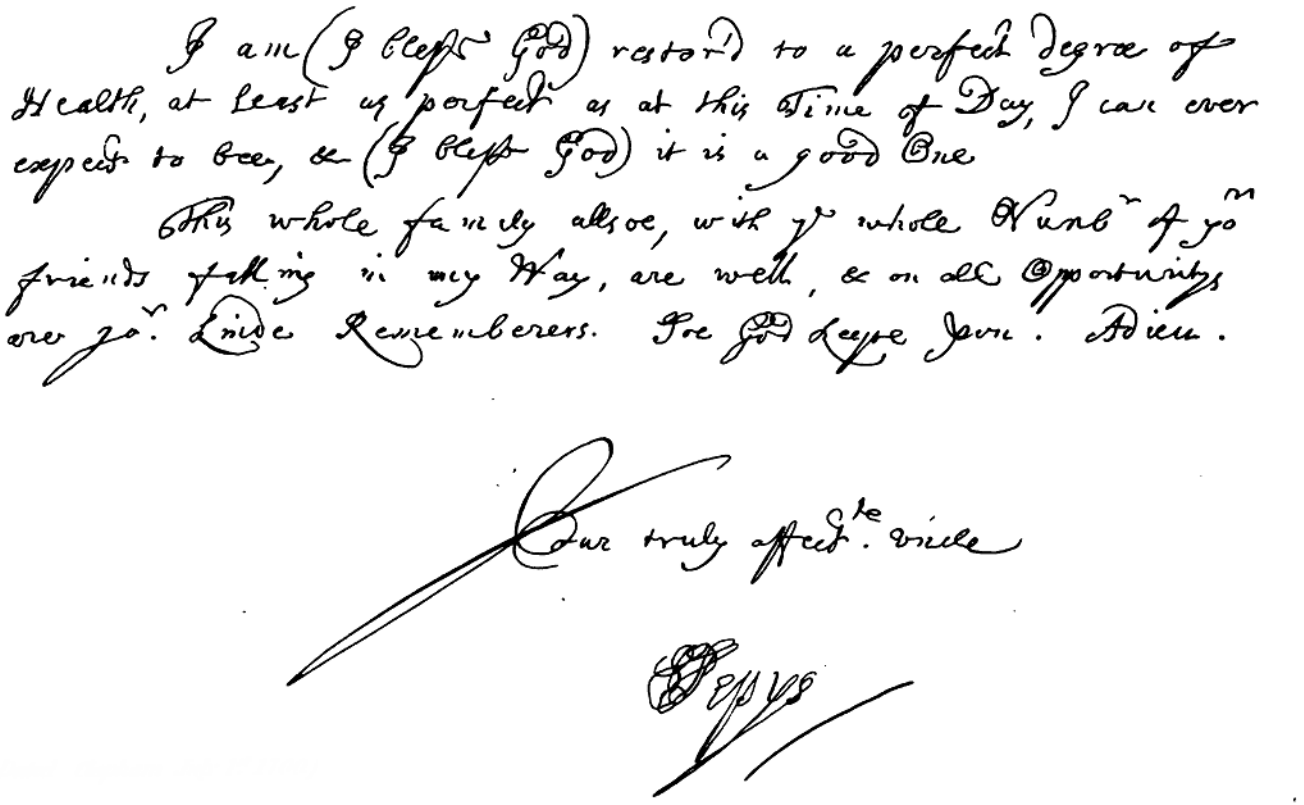


Statistical significance

- **Wilcoxon rank-sum test** (Wilcoxon 1945; Mann & Whitney 1947)
  - Order samples based on normalised frequency, see how surprising the result is
- 1000s of hypotheses tested → **false discovery rate control** (Benjamini & Hochberg 1995) at FDR = 0.1
  - ➢ Significance threshold $p < 0.0035$

- Thanks to Jukka Suomela for methodological assistance!

Results

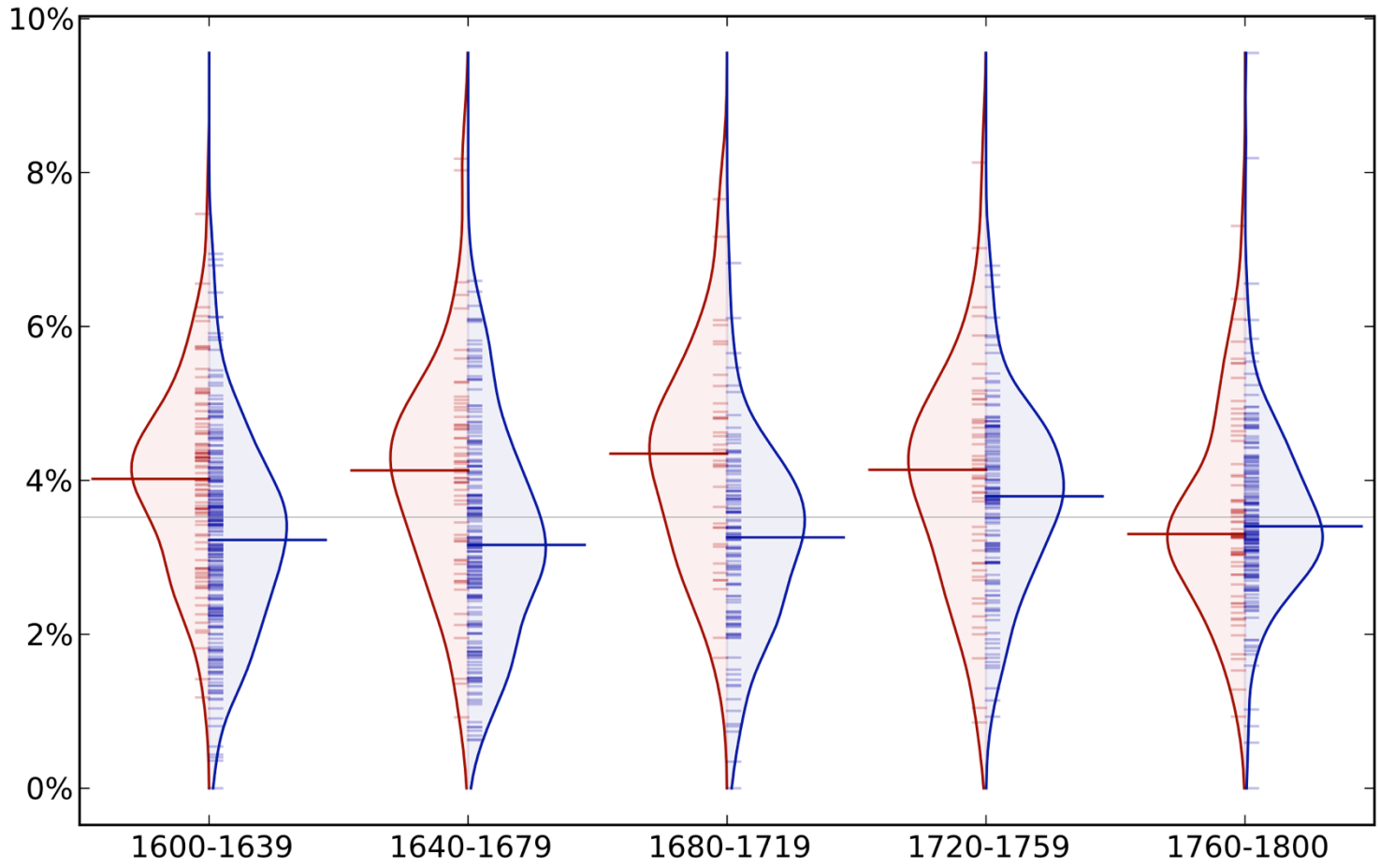

I am (I bless God) restor'd to a perfect degree of Health, at least as perfect as at this Time of Day, I can ever expect to bee, & (I bless God) it is a good One

This whole family allsoe, with ye whole Numbr of yor friends falling in my Way, are well, & on all Opportunitys are yor. Linde Rememberers. Soe God Keepe Jou. Adieu.

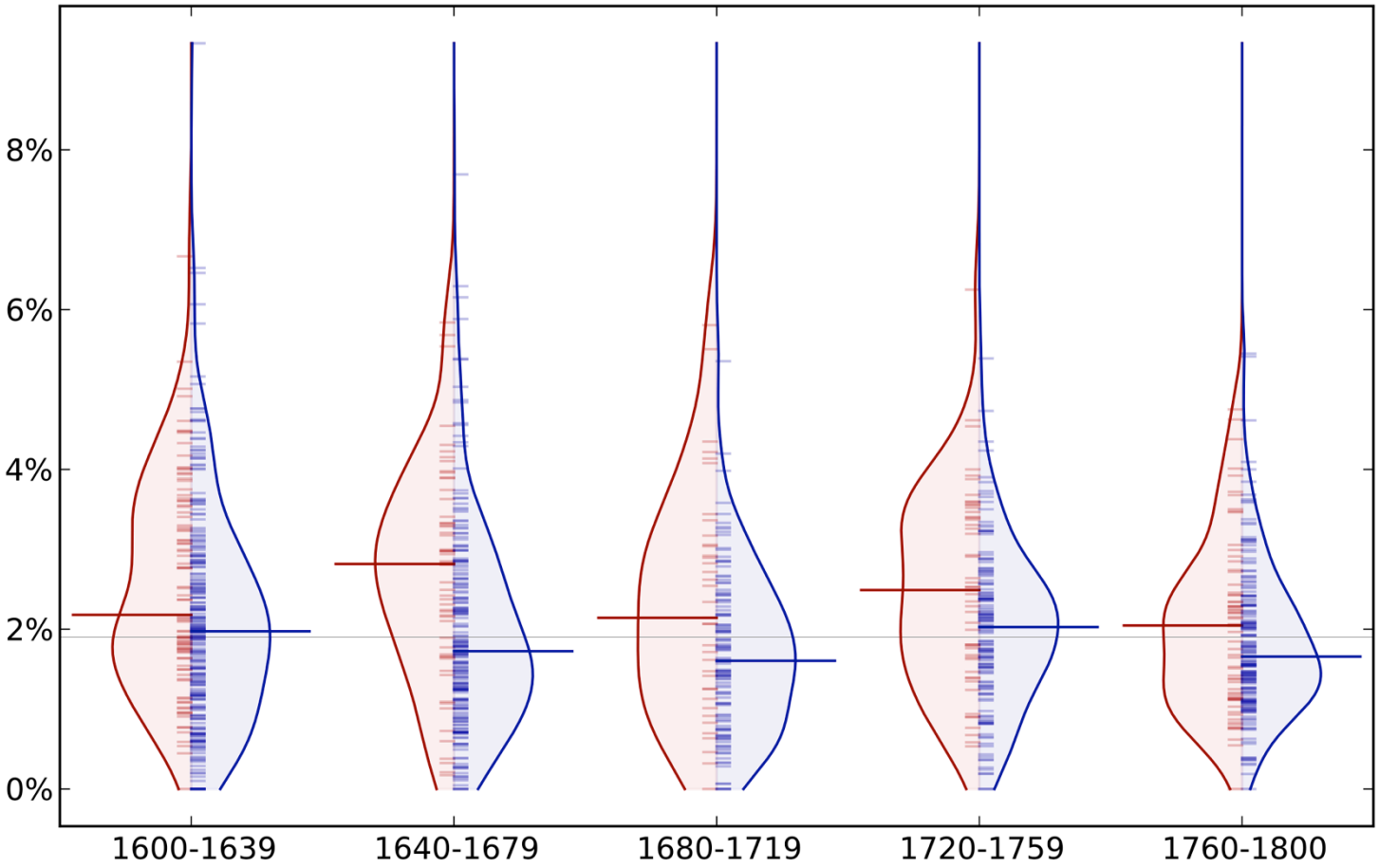
Your truly affect.te uncle

Pepys

Overall gender differences: *I*



Overall gender differences: *you*



Overall gender differences

- Gender affects pronoun use
- Women use more *I* and *you* than men
- **BUT**
  - The difference is not significant in every period
  - Variation across time (a levelling tendency in the 18th century)
  - Women: more FN letters, men: more T letters

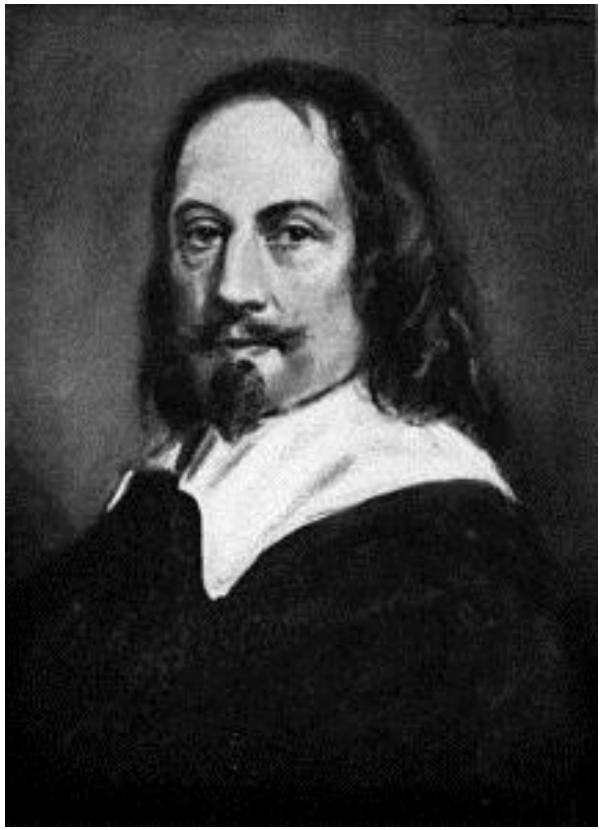
Results

Pronoun use within the nuclear family

- Letters written by men
  1. Son – father
  2. Father – son
  3. Husband – wife

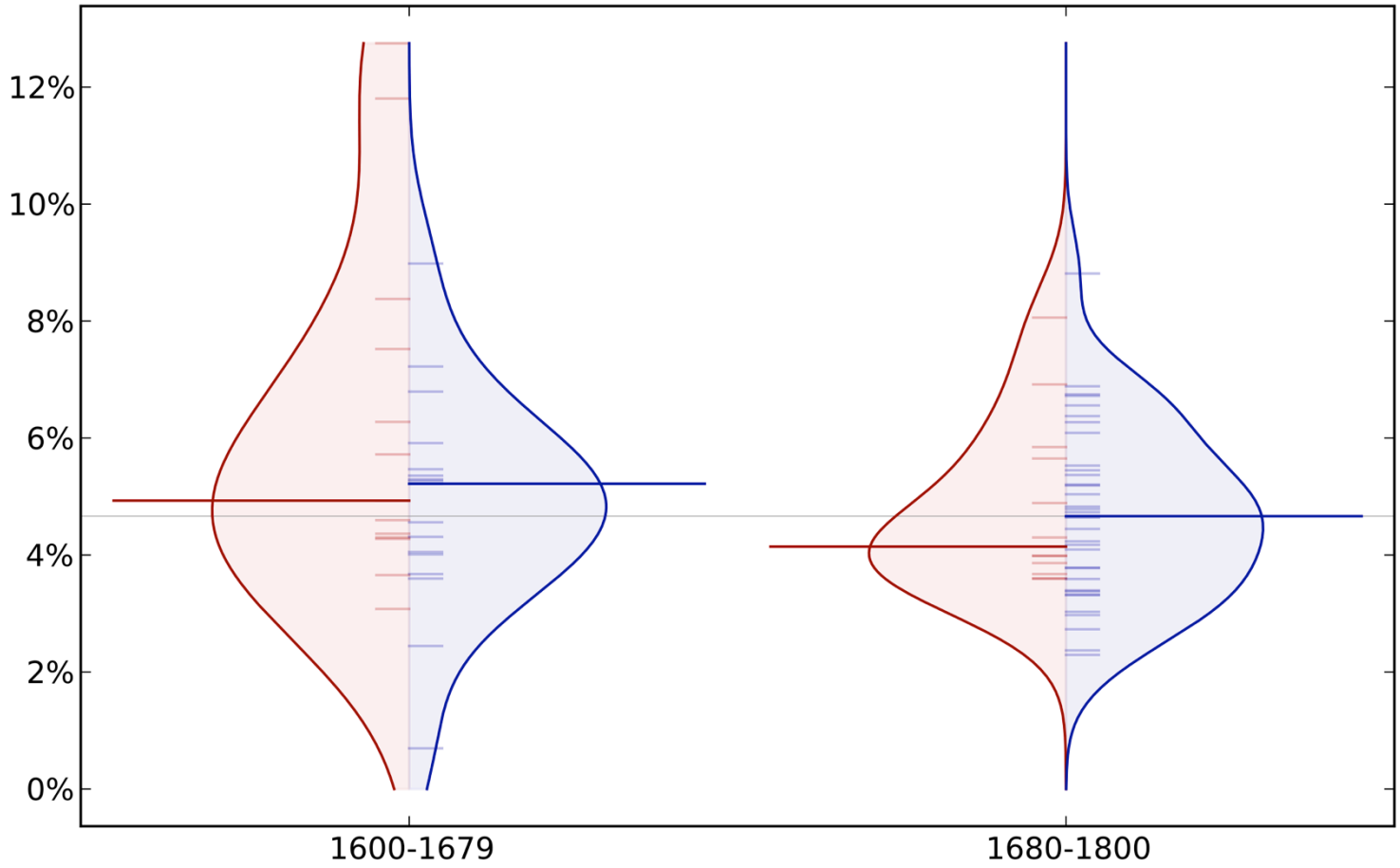
- Letters written by women
  1. Wife – husband
  2. Mother – son

Case 1: Fathers vs. sons



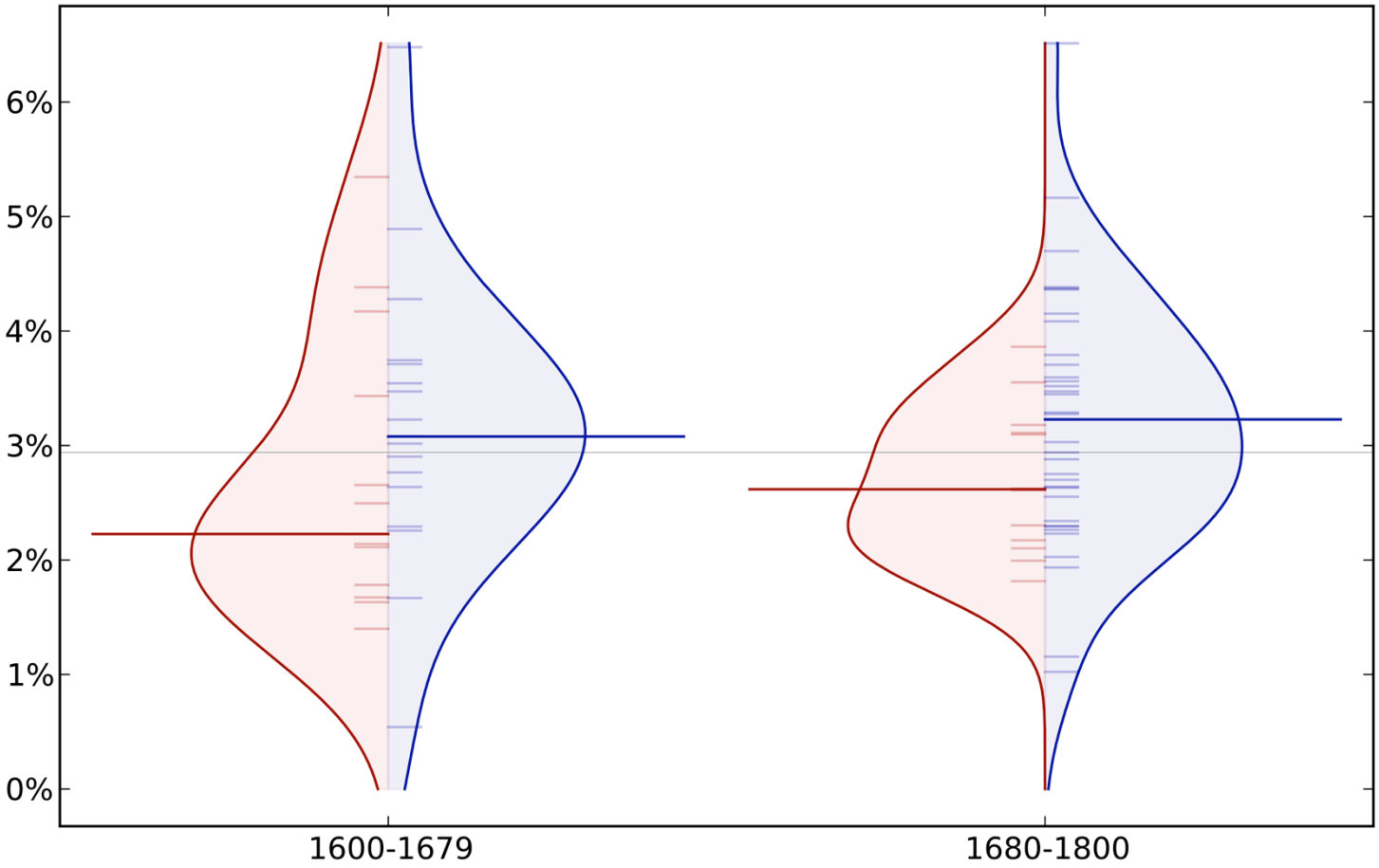
Results

Father – son vs. son – father: *I* + *you*



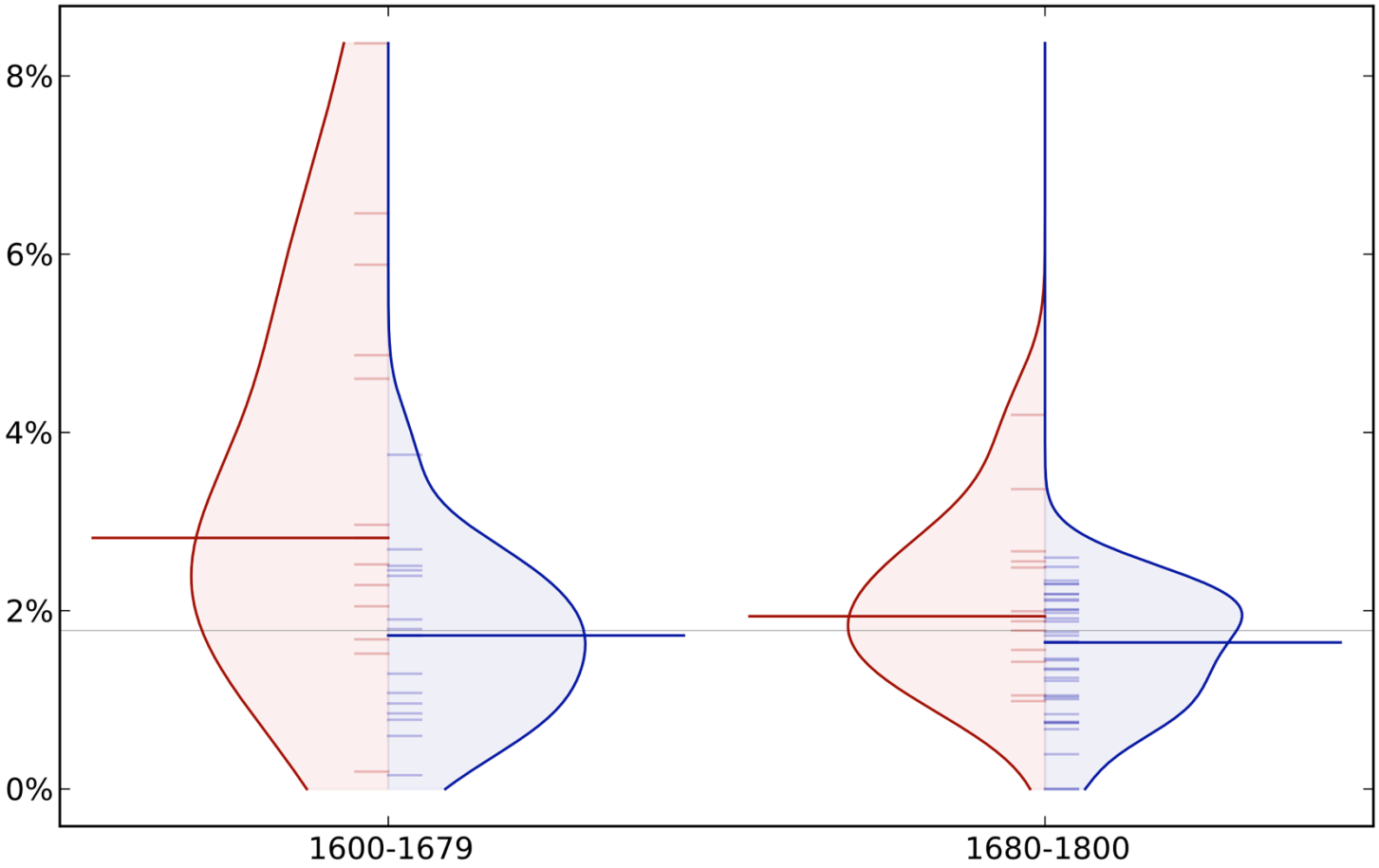
Results

Father – son vs. son – father: *I*



Results

Father – son vs. son – father: *you*

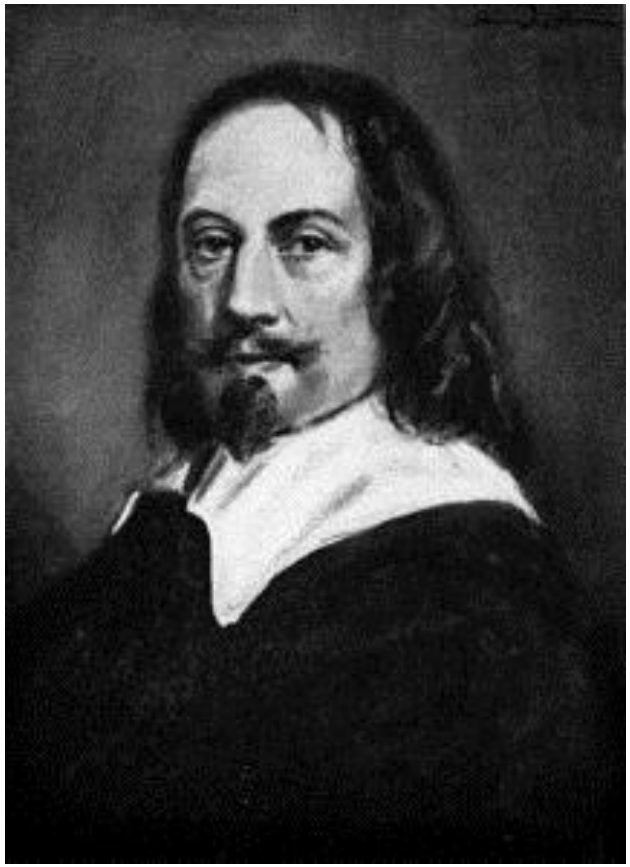


Sons use *I* more than fathers

My Lord, That dutie which inioynes this taske, is not more the same then the matter which **I** have to write of. In that method of exercises, and studie, which **I** have heretofore made known unto your Lordship doe **I** stil goe on: though in the successe, **you** shall perhaps meet with little that may take **you**, yet **I** hope in my desires **you** will find something at least capable of your pardon. **I** am drawing nigh to Fortifications, but such as imagination is able to raise, and to demolish: if honor were not so strictly espoused to danger, **I** could wish your Lordship had at this time no harder a taske in hand, that **I** might boldly presume of your safety…

(Edward Conway, aged 17, to his father, 1640)

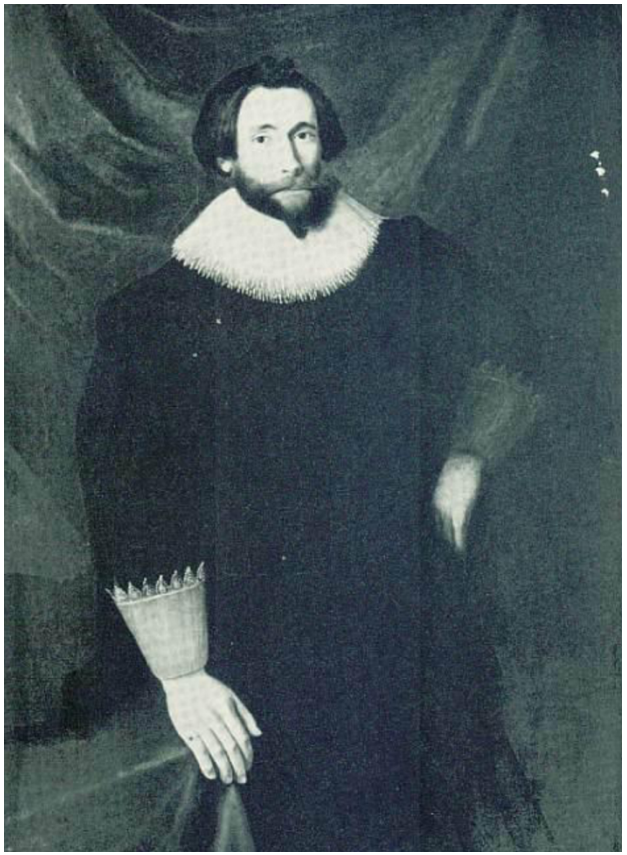
Fathers use *you* more than sons



Honest Tom, God blesse & protect thee & mercifully lead **you** through the wayes of his providence. **I** am much greived **you** have such a cold, sharpe & hard introduction, wch addes newe feares unto mee for your health, whereof pray bee carefull & as good an husband as possible, wch will gayne **you** credit & make **you** better trusted in all affayres. **I** am sorry **you** went unprovided with bookes without which **you** cannot well spend time in those great shipps. If **you** have a globe **you** may easily learne the starres as also by bookes. Waggoner **you** will not bee without wch will teach the particular coasts, depths of roades & how the land riseth upon severall poynts of the compasse. Observe the variation of the compasse. Blundevill or Moxon will teach **you** severall things.

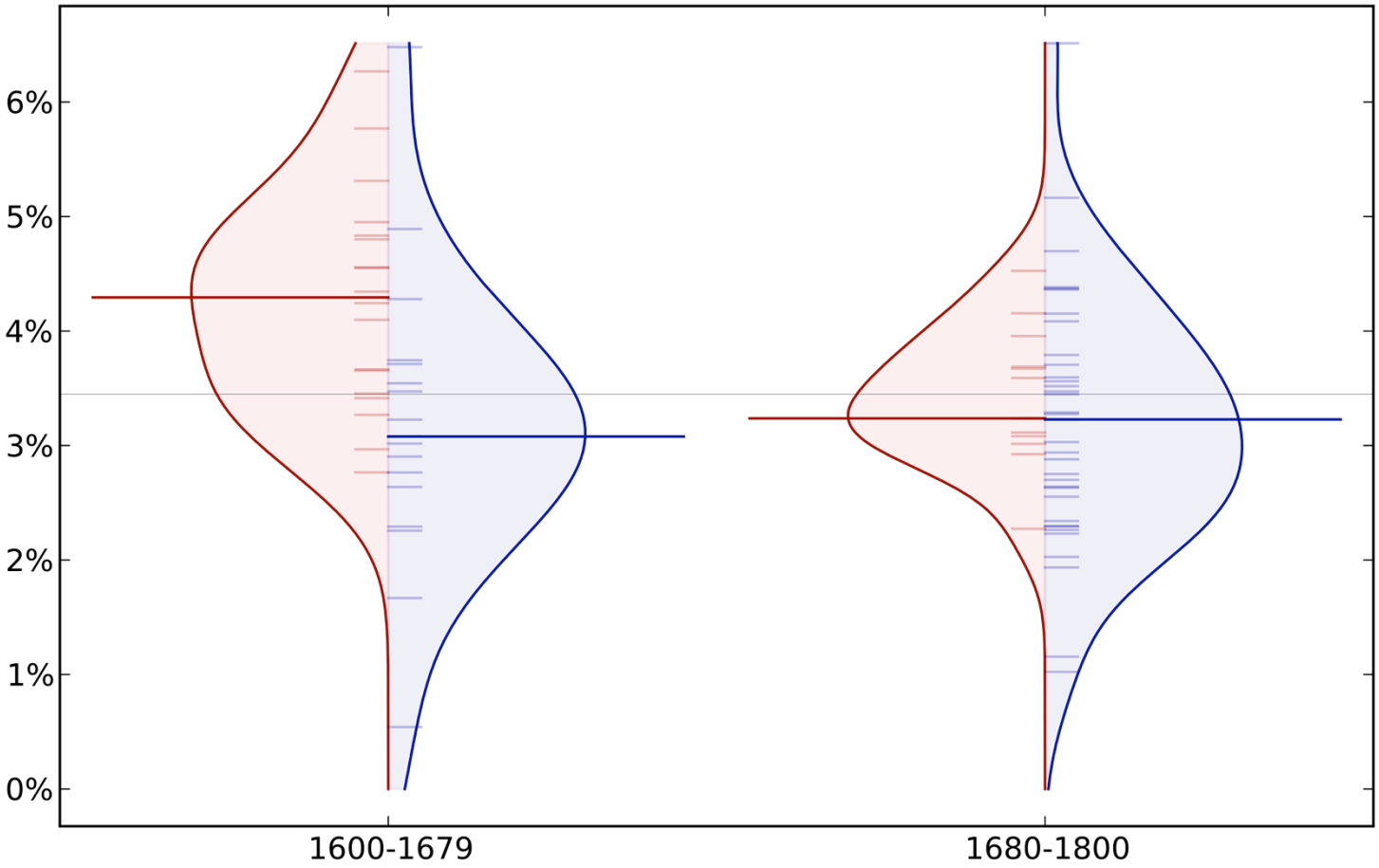
(Sir Thomas Browne, aged 60, to his son, 1665)

Case 2: Husbands vs. sons



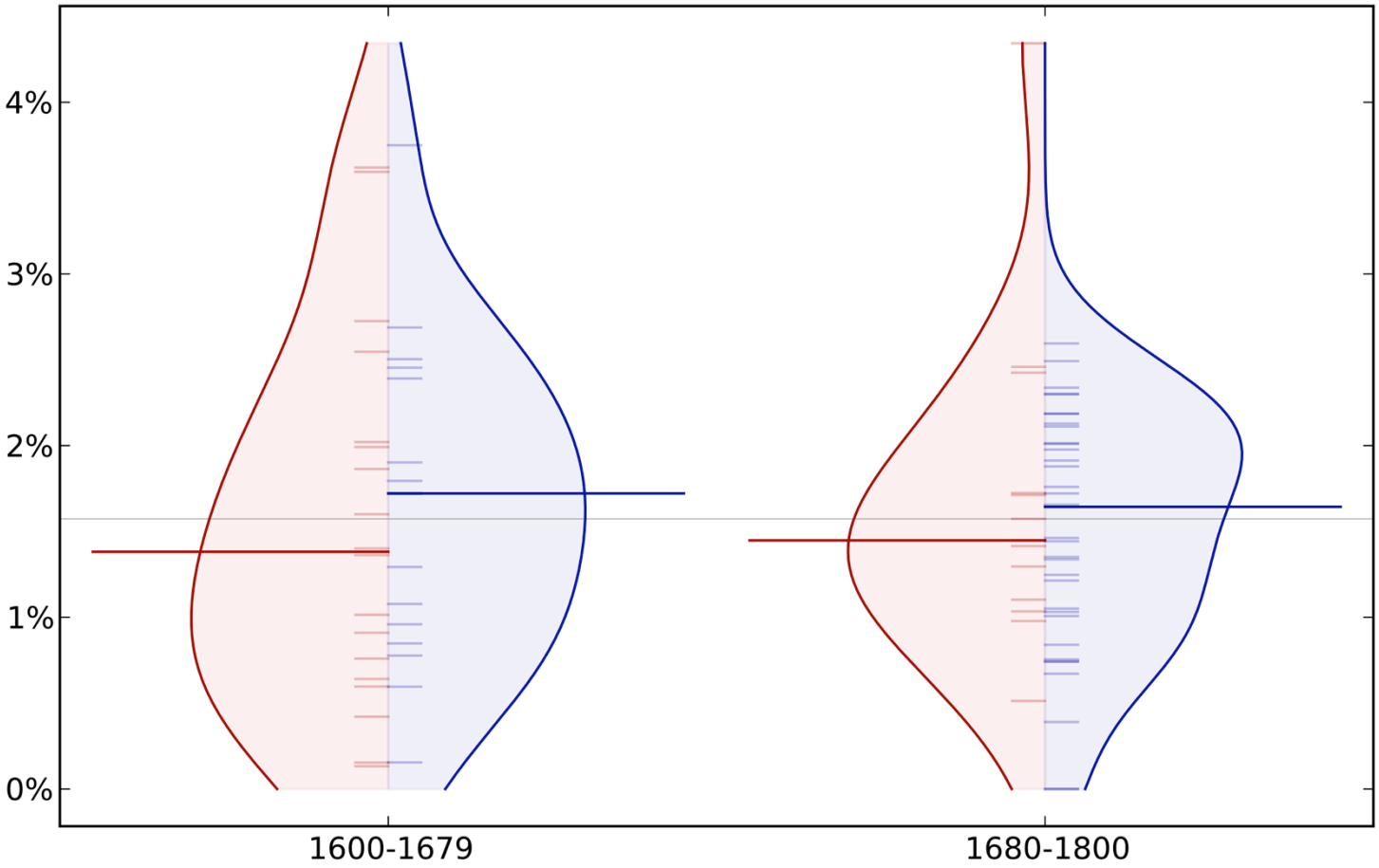
Results

Husband – wife vs. son – father: *I*

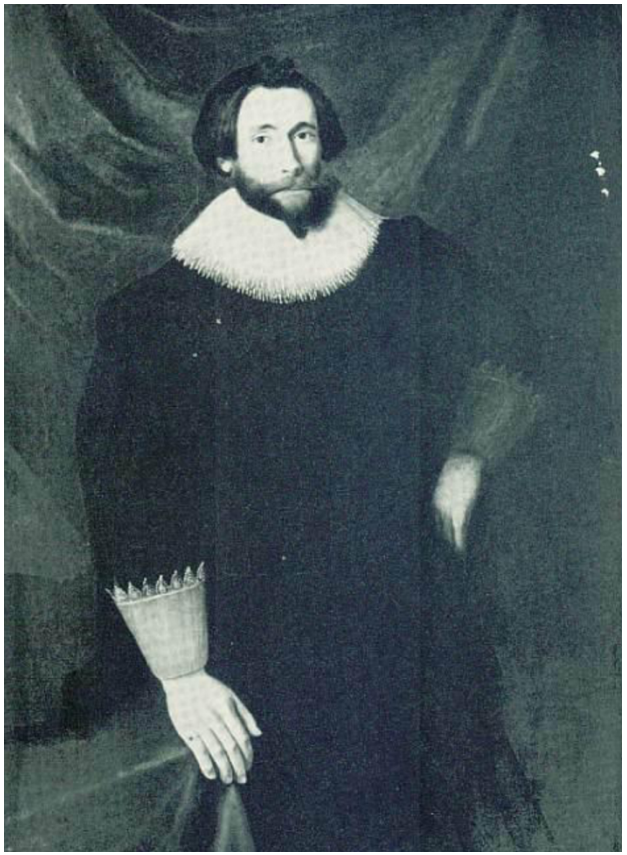


Results

Husband – wife vs. son – father: *you*



Husbands use *I* even more and *you* even less than sons



My busines with Jones falls out still vntowardly, ther being a compacte of knavery amongst them to defraud the creditors but I hope to prevaile for my mony notwithstanding in a short time. I haue made inquiry of the land of Starbourg, and haue spoken with Beareblock, which M^{r}. Seriant Richardson did take very ill, fearing least I had made an agreement with him for his statute, which I would to god I had. I cannot tell yett how we shall agree but as neere as I can I will deale to our best advantage.

(Thomas Knyvett, aged 26, to his wife, 1622)

Husbands, sons and fathers

- Implicational scale for *I*

Husbands	>	Sons	>	Fathers

- Implicational scale for *you*

Fathers	>	Sons	>	Husbands

- Statistical significance only for husbands vs. fathers, 1600–1679 (both *I* and *you*)

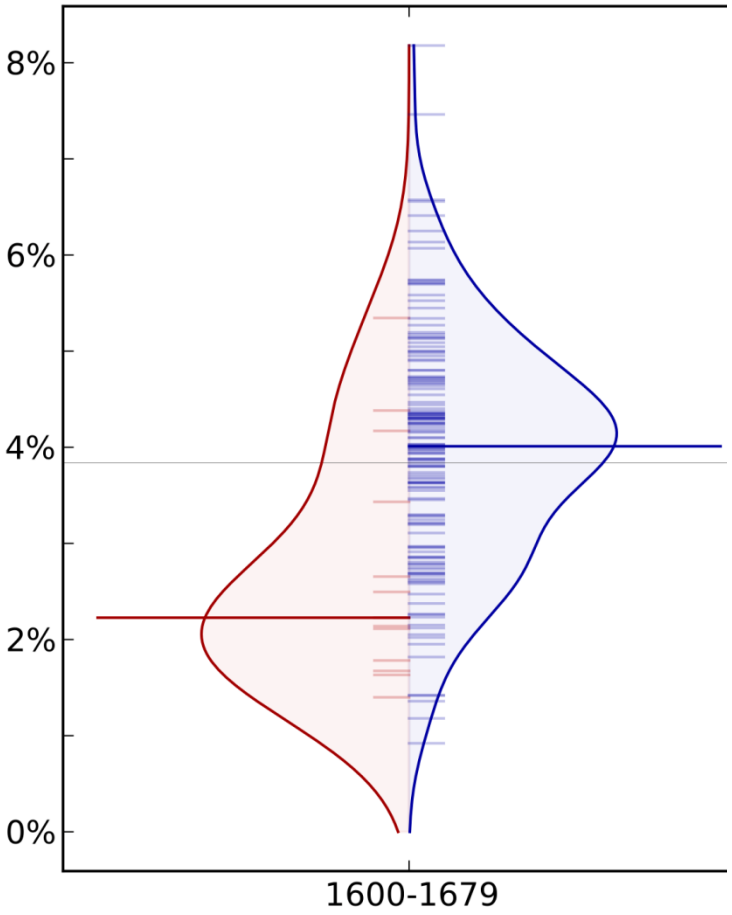
Family letters in CEEC

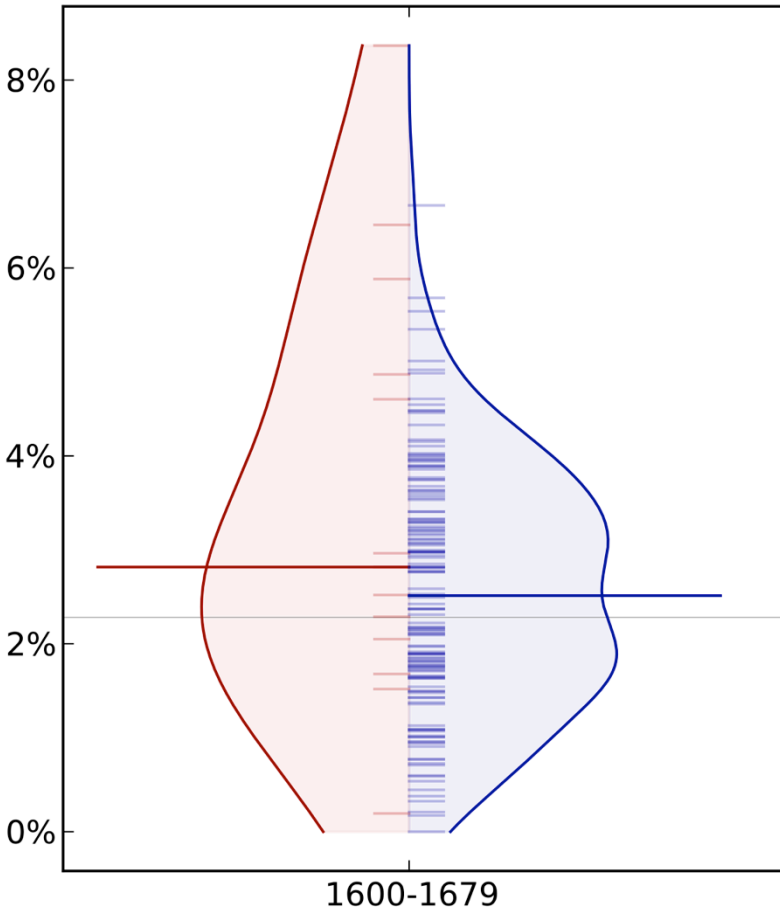
	1600–1679		1680–1800		Entire period	
husband > wife	101,015	40%	15,716	7%	116,731	24%
wife > husband	31,502	12%	51,910	22%	83,412	17%
father > son	26,756	10%	39,560	17%	66,316	13%
son > father	29,090	11%	94,931	40%	124,021	25%
mother > son	41,734	16%	18,902	8%	60,636	12%
son > mother	25,039	10%	15,381	7%	40,420	8%
Total	255,136		236,400		491,536	

Importance of relationship categories to corpus design

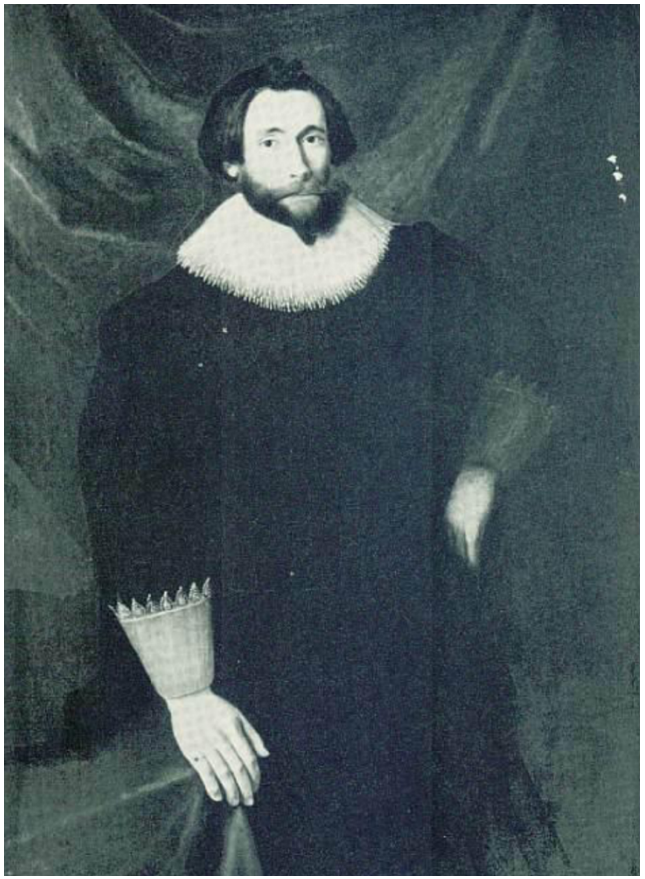
Fathers vs. women

Father – son vs. all women: *I* and *you* (1600–1679)

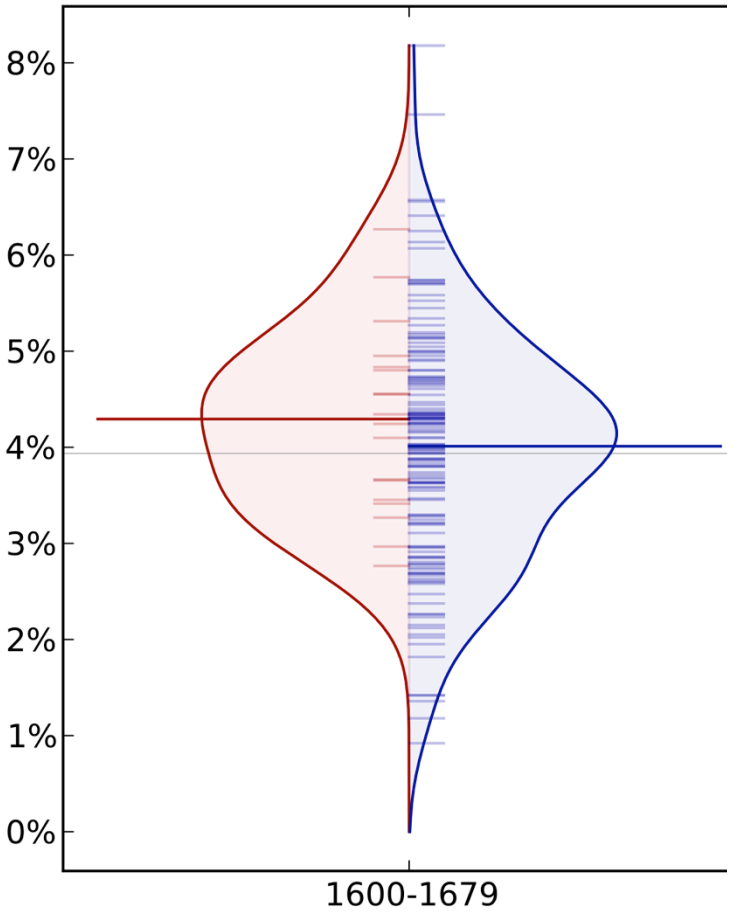


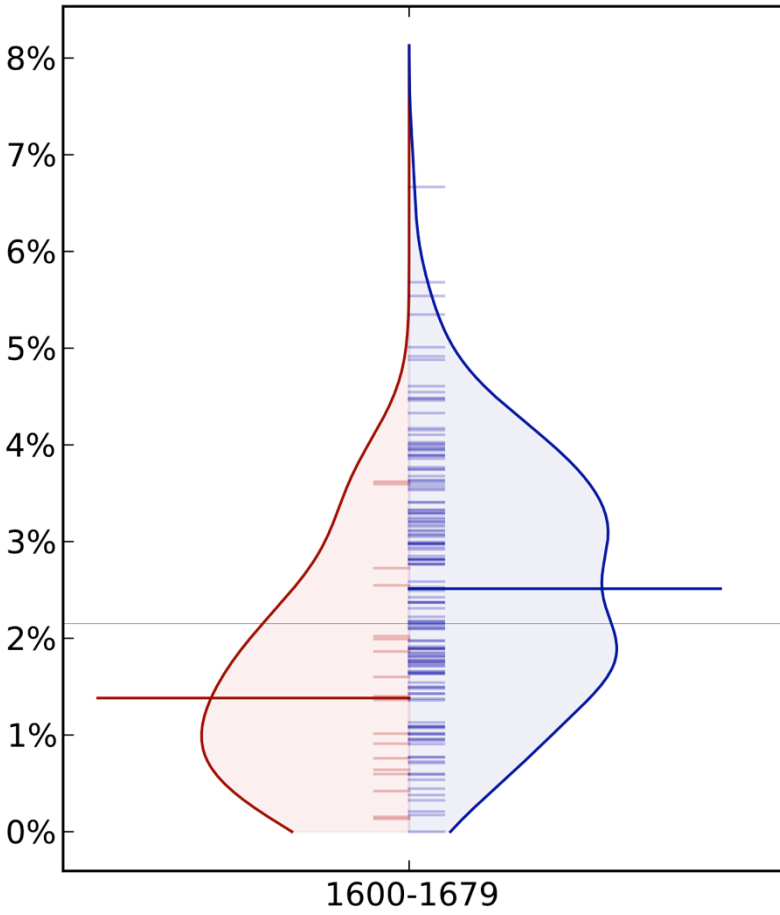


Husbands vs. women



Husband – wife vs. all women: *I* and *you* (1600–1679)





Gender differences within the nuclear family

Arthur Capel, 1st Earl of Essex with his wife, Elizabeth, Countess of Essex

Gender differences within the nuclear family

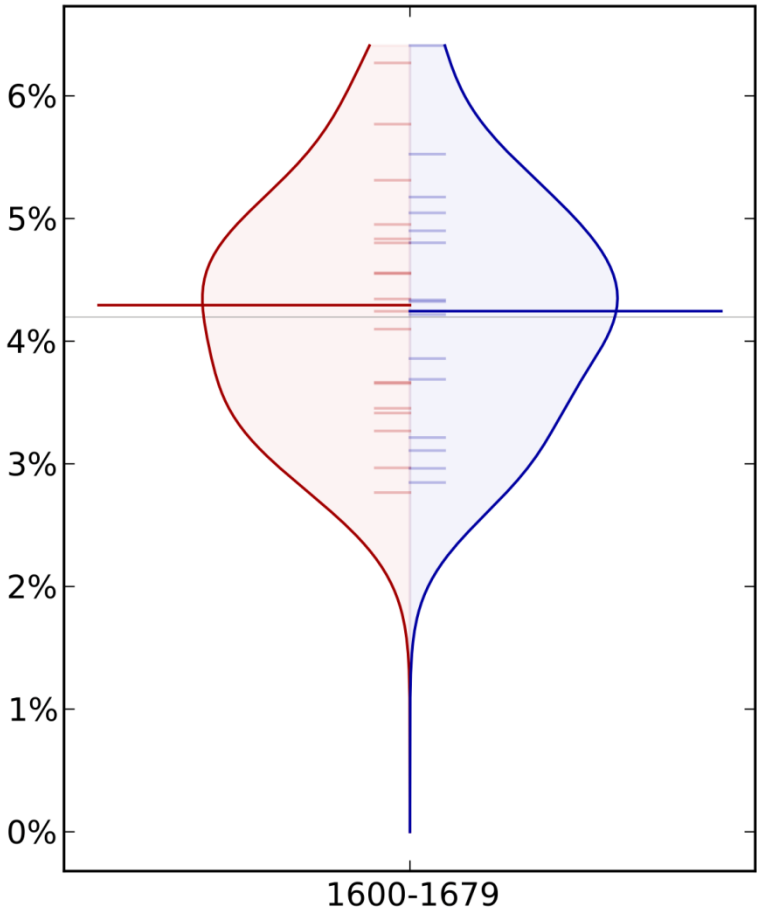
Sender-recipient	Sender-recipient	Period	Pro	+/-	*p* value
Mother-son	Father-son	1600–1679	I	+/-	0.00286
Wife-husband	Husband-wife	1680–1800	I	+/-	0.00226
Wife-husband	Husband-wife	1600–1679	You	+/-	0.00040
Mother-son	Son-mother	1600–1679	You	+/-	0.00005

Gender differences within the nuclear family

- The results point to women's overuse of pronouns
- Most of the differences observed in our study are strong tendencies
  - Statistically significant results only obtained when women use more pronouns than men
- Husband–wife letters (1600–1679)
  - The only category where men use more *I* than women
  - Not statistically significant

Results

Husbands vs. wives: *I* (1600–1679)



Conclusion

Gender-based or relationship-based?

- Sender–recipient relationship is clearly important
- Gender is a more important factor
  - In inter-gender comparisons (e.g. husband–wife vs. wife–husband) women consistently used more pronouns than men
- Relationship categories should be taken into account in both research and corpus design

Thank you!

References

Argamon, S., M. Koppel, J. Fine & A. R. Shimoni. 2003. “Gender, genre, and writing style in formal written texts”. *Text* 23(3): 321–346.

Benjamini, Y. & Y. Hochberg. 1995. “Controlling the false discovery rate: A practical and powerful approach to multiple testing”. *Journal of the Royal Statistical Society: Series B (Methodological)* 57(1): 289–300.

Biber, D. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, D. 1995. *Dimensions of Register Variation: A Cross-Linguistic Comparison.* Cambridge: Cambridge University Press.

Biber, D. & J. Burges. 2000. “Historical change in the language use of women and men: Gender differences in dramatic dialogue”. *Journal of English Linguistics* 28(1): 21–37.

CEEC = *Corpus of Early English Correspondence*. 1998. Compiled by T. Nevalainen, H. Raumolin-Brunberg, J. Keränen, M. Nevala, A. Nurmi and M. Palander-Collin at the Department of Modern Languages, University of Helsinki. http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/

CEECE = *Corpus of Early English Correspondence Extension*. Compiled by T. Nevalainen, H. Raumolin-Brunberg, S. Kaislaniemi, M. Laitinen, M. Nevala, A. Nurmi, M. Palander-Collin, T. Säily and A. Sairio at the Department of Modern Languages, University of Helsinki.

CEECSU = *Corpus of Early English Correspondence Supplement*. Compiled by T. Nevalainen, H. Raumolin-Brunberg, S. Kaislaniemi, M. Laitinen, M. Nevala, A. Nurmi, M. Palander-Collin, T. Säily and A. Sairio at the Department of Modern Languages, University of Helsinki.

Kampstra, P. 2008. “Beanplot: A boxplot alternative for visual comparison of distributions”. *Journal of Statistical Software* 28: Code Snippet 1.

Mann, H. B. & D. R. Whitney. 1947. “On a test of whether one of two random variables is stochastically larger than the other”. *Annals of Mathematical Statistics* 18(1): 50–60.

Palander-Collin, M. 1999. “Male and female styles in seventeenth-century correspondence: I THINK”. *Language Variation and Change* 11: 123–141.

Rayson, P., G. Leech & M. Hodges. 1997. “Social differentiation in the use of English vocabulary: Some analyses of the conversational component of the British National Corpus”. *International Journal of Corpus Linguistics* 2(1): 133–152.

Säily, T., T. Nevalainen & H. Siirtola. 2011. “Variation in noun and pronoun frequencies in a sociohistorical corpus of English”. *Literary and Linguistic Computing* 26(2): 167–188.

Vartiainen, T., T. Säily & M. Hakala. Forthcoming. “Variation in pronoun frequencies in early English letters: Gender-based or relationship-based?” In J. Tyrkkö, Olga Timofeeva & Maria Salenius (eds.). *Ex Philololgia Lux: Essays in Honour of Leena Kahlas-Tarkka*. Helsinki: Modern Language Society.

Wilcoxon, F. 1945. “Individual comparisons by ranking methods”. *Biometrics Bulletin* 1(6): 80–83.