

Changes in Uniqueness of Chinese Given Names Over Time

Tyler Cheung, Nikita Daga, Kareena Legare, Shelby Tisdale,
Luke Vrotsos, Ken Ye

Abstract

In Chinese culture, names are chosen with great care and are very meaningful. Previous research has indicated that names may also contribute to the way a person is perceived, makes decisions, and experiences life, with the uniqueness of a name playing a role in these outcomes. This paper aims to explore what factors contribute to the uniqueness of a name, and more specifically, how the uniqueness of a character changes over time. The parts per million (ppm) of characters in Chinese given names is used to indicate how unique a character is. This study presents visualization-based analysis to answer the research question. We found that the uniqueness of Chinese characters has increased over time, and that the average ppm of characters with the highest perceived valence, warmth, and competence has also increased. Our findings align with the cultural shift in China toward individuality due to the rise in globalization. This departure from the traditional emphasis on fulfilling family wishes during the naming process represents a societal transformation and an evolution in the significance of modern names.

Introduction

One's first name wields social and psychological influence. A range of research regarding the influence of first names in alphabet-based languages like English has revealed that various characteristics, such as perceived positivity, warmth, gender association, trustworthiness, and competence, may be associated with certain names. Thus, a first name with a particular reputation may bias others' perceptions of the name holder. Less studied is the influence of given names in non-alphabet-based languages such as Chinese, in which names are represented pictographically (Du et al. 2021). A small body of research explores Chinese naming trends but less of this research examines how Chinese naming practices and characteristics associated with given names have changed over time. Evidence points to significant changes in Chinese culture

over the past century, but research has yet to bridge these cultural shifts with characteristics of Chinese given names.

In Chinese culture, a person’s surname is usually inherited from the paternal side. Given names, which typically consist of 1-4 characters, however, are chosen with great care. According to a 1927 paper by Ching-Chao Wu, some parents choose names that express the wish of the family – for example, Pei Yuan, which means “educating descendants,” or Shih Tsu, which means “perpetuation of the line.” Others choose names which connote desirable qualities – for example, Chien Ho, which means “goodness and peace” (Wu 1927).

Additionally, research by Chua has shown the evolution of Han Chinese names and how these names have served purposes beyond mere identification. Popular Chinese names have changed from the 1940s to 2008, encoding the historical happenings in the country and the dynamic desirable qualities valued by the people. Before the 1960s, popular Chinese names conveyed ideas of reconstruction. By the 1970s, the success of the Chinese emperors was reflected in male baby names with meanings such as “building the nation” and “building the army”. By the early 21st century, names meaning “space travel” grew in popularity, resembling the technological and economic success of the country. The research also showed the prominent characteristics expected of sons and daughters. Typical masculine traits were “ambitious”, “forthright”, and “cultural”, while feminine traits included “beautiful” and “pristine” (“What Can We Tell from the Evolution of Han Chinese Names?” 2021).

Despite the continued prevalence of these traditional naming practices, Cai et al’s paper indicates an increasing emphasis on need for uniqueness (NFU), which can be understood as a positive striving for differentness relative to other people in China, a society traditionally oriented toward collectivism. While Cai et al.’s first study revealed a notable rise in self-reported NFU among Chinese individuals across different generations, the second study used a different proxy, which is especially relevant to our research. The researchers analyzed the names given to newborn babies in China over the past 50 years and observed a growing trend of parents using unique characters to name their children. This highlights how name uniqueness plays a key role in naming practices as well as name popularity (Cai et al. 2018).

Furthermore, a study by Du, Gu, and Dong et al. sought to explore how perceptions of warmth and competence are associated with recognizability in characters used in Chinese given names. Researchers concluded that those with names considered easily recognizable were generally perceived as warmer but not more competent than those with unique names, independent of pronunciations and semantic meanings of names. Both warmth and competence ratings, however, contributed to judgements of one’s trustworthiness. Despite increased prevalence of unique characters in Chinese given names in recent centuries, the researchers suggest that rare Chinese characters should be avoided in name selection due to adverse effects on warmth perception and trust judgements. This study motivates further exploration of changes in warmth, competence, and uniqueness of characters in Chinese given names over time (Du et al. 2021).

Much of the existing literature examines the influence of given names in Chinese culture but lacks information on how Chinese given names have changed over time. Furthermore, prior research that investigates changes in naming trends over time typically lacks insight into how different names are perceived. Additionally, we can find little literature pointing to which factors influence character popularity in Chinese given names.

In this study, we will utilize data visualizations to conduct a visual exploration of how the uniqueness of Chinese given names has evolved from the 1930s to 2008. By analyzing determining factors of character popularity in Chinese given names, we aim to gain insights into changing cultural and societal trends in China. Specifically, we will explore the relationships between the frequency of character usage and name attributes such as warmth, competence, and valence. Moreover, we examine the change in parts per million (ppm) of Chinese characters between decades in the time period of interest, providing an indicator of name uniqueness and characters used in names. Through our visual analysis, we seek to highlight the shifting significance and uniqueness of modern Chinese names and the cultural transformation accompanying

Data

The data we are using is from an R package compiled by Dr. Bruce Bao that contains information about 1,806 Chinese surnames and 2,614 Chinese characters used in Chinese first names from 1930 to 2008. It was originally obtained from the National Citizen Identity Information Center (NCIIC) of China in 2008. There are five datasets within the package, and the dataset chosen for this research report contains information about the characters used in Chinese given names. Our report intends to analyze how Chinese name popularity has changed over time with respect to various factors affecting the name. The given names dataset was the only dataset with information about the change in name frequency over time and the corresponding warmth, valence, competence, and gender ratings of the names.

The main variables in this dataset are summarized in the table below. The variables for the time frame 1970-1979 are used as an example and variables exist for decades from 1930 to 2008 in the dataset, with all decades before 1960 grouped into a single variable.

Variable	Class	Description
character	character	A single character Chinese name
bihua	integer	The number of strokes required to write the character
name.gender	double	How masculine or feminine the name is
n.1970_1979	integer	Frequency of the name in the given time frame
name.uniqueness	double	The frequency of the name per million people in the time
corpus.ppm	double	Numerical value of how unique the name is

Variable	Class	Description
corpus.uniqueness	double	The frequency of the character per million Chinese characters used in the given time frame
name.valence	double	Numerical value of how unique the character
name.warmth	double	Subjective numerical rating of how positive the name is
name.competence	double	Subjective numerical rating of how likely someone with the character is likely to have competence traits
ppm.1970_1979	double	The frequency of the name per million people in the time frame

Name warmth and name competence are subjective ratings, calculated by asking 10 Chinese raters – 5 male and 5 female – on how likely someone with a given Chinese character in their name is likely to have these traits on a Likert scale of 1 to 5. A higher rating indicates a higher likeliness. Name valence is computed on a similar scale with 16 raters – 9 male and 7 female – on how positive the name is with higher ratings indicating greater positivity.

Name gender values are computed on a scale of -1 to 1 with -1 indicating a completely feminine name, 1 indicating a completely masculine name and 0 being a gender-neutral name.

Name Uniqueness and Character Corpus Uniqueness are both calculated on a scale of 1 to 6 using the formula $-\log_{10}(P_{\text{char}} + 10^{-6})$ where P_{char} for Name Uniqueness represents the percentage of a character used in all single-character and multi-character given names in the Han Chinese population within a specific birth year and P_{char} for Character Corpus Uniqueness represents the percentage of a character in a contemporary Chinese corpus (Bao 2021).

Methods

The objective of our analysis was to identify the determining factors of the popularity of a character usage in Chinese people’s given names over time. All analysis was performed using R. We found that visualizations were more effective in answering our research question compared to linear regression models. Our visualizations primarily focused on time series analysis of the warmth, valence, competence, and ppm of Chinese characters.

The original dataset contained six variables for the parts per million (ppm) of a character during six time periods. In order to make this data more usable, we pivoted the data frame along these columns to create two new variables (ppm and decade). This gave us six rows for each character, one for each time period, and allowed us to better perform our analysis.

Our first set of visualizations consists of scatter plots that illustrate the relationship between the frequency and the warmth, competence, and valence of Chinese characters over time. To

create the first plot, we selected 400 characters from the pivoted data, specifically the top and bottom 200 characters based on their warmth ratings. In order to reduce overlap, a small amount of random variation was introduced using the “geom_jitter” function in the ggplot package. Moreover, the “parts per million” variable was log transformed to reduce the skewness in the data, enabling the detection of more subtle patterns. The points on the scatter plot were colored according to their warmth ratings. This process was then repeated, replacing the warmth variable with competence and valence. By employing these visualizations, we aimed to gain a deeper understanding of the relationship between the variables of interest (warmth, competence, valence), their association with the parts per million (ppm) metric, and how these patterns changed throughout time.

The second set of visualizations examines the change of the summed ppm of Chinese characters per decade over time. Using the pivoted data, we selected the top 100 most common characters used in Chinese names during each time period. Their ppms were added together to find the total ppm that these characters make up. These values were plotted on a line graph to show how the summed ppms changed throughout time. The goal of this plot was to examine the change in uniqueness of characters over time in Chinese names. Presumably, the more unique characters are, the lower their combined ppm will be.

Results

In Figure 1, we examined how the natural logarithm of the ppm of the 400 selected characters changed over time, and whether warmth was related to this. The first visible trend is the decrease in the ppm of the warmest characters and an increase in the ppm of the least warm characters throughout time. It also appears that the range of overlap of the ppm between the warmer and less warm characters (indicated by point color) is decreasing over time. This is primarily caused by a decrease in the range and increase in the average of ppm values for the warmest characters.¹

Figure 2, examining the competence of characters over time, shows similar patterns. Although it does not appear that the characters perceived as most competent are decreasing in ppm, there is an increase throughout the time periods in ppm among the characters perceived as least competent. Additionally, there is another visible, and much stronger, polarization between the ppms of the most and least competent characters over time, mostly due to the decreased range and increased average of ppm values for the most competent characters.²

The third plot in this set of visualization, Figure 3, once again shows similar patterns to those seen above. The characters receiving the highest ratings for valence are decreasing in ppm throughout time, and those receiving the lowest valence ratings are increasing. The same grouping is also visible in this plot, and for the same reason, marked by the increasingly clear

¹See Figure 6 in appendix for correlation matrix between per-decade ppm and competence.

²See Figure 7 in appendix for correlation matrix between per-decade ppm and valence.

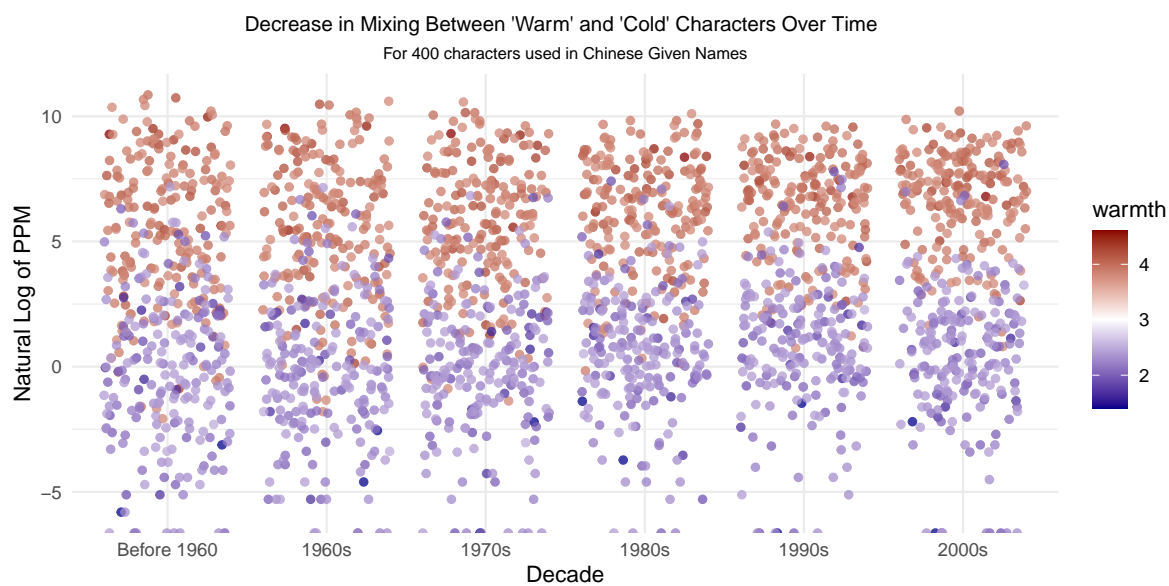


Figure 1: Frequency and Warmth of Chinese Characters Over Time

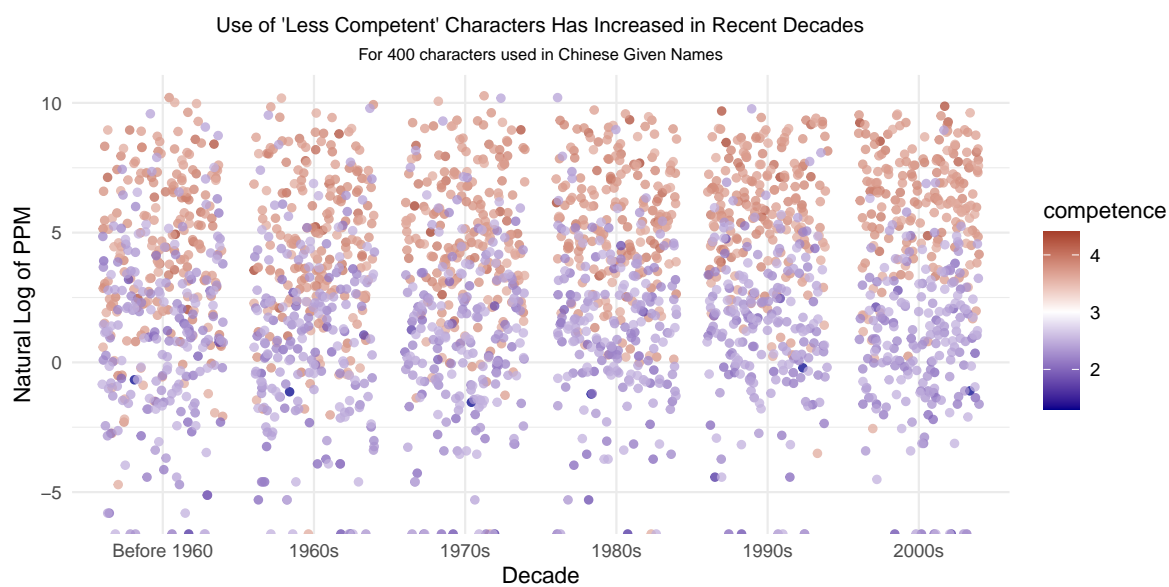


Figure 2: Frequency and Competence of Chinese Characters Over Time

separation of point colors from decade to decade.³ The points on this plot are noticeably darker than those of previous plots, indicating that reviewers were more likely to give names a strongly negative or positive score of valence than warmth or competence.

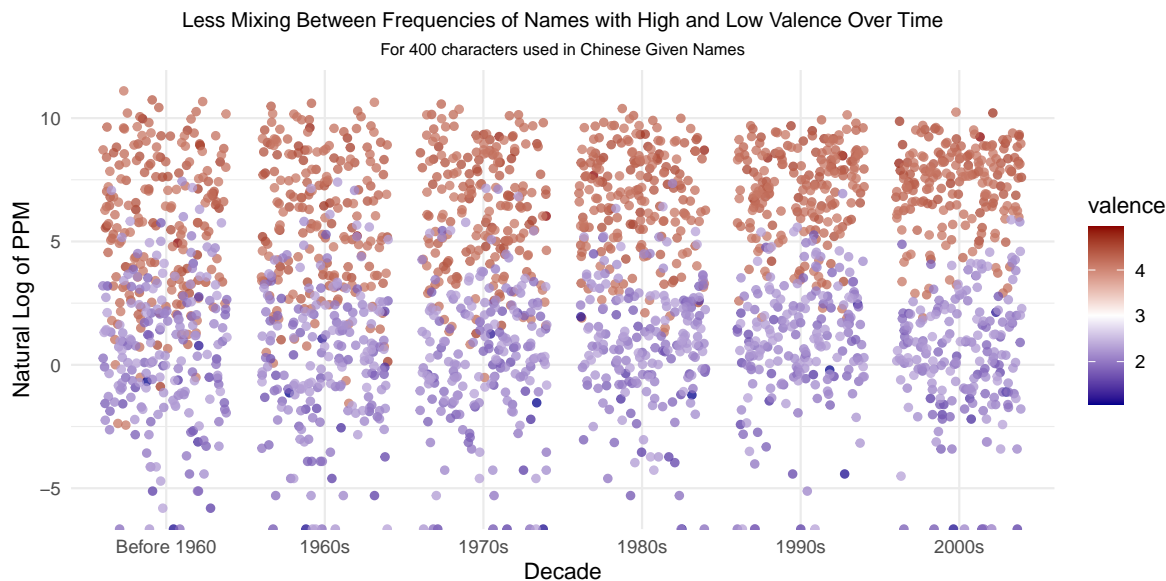


Figure 3: Frequency and Valence of Chinese Characters Over Time

The grouping pattern of points seen in all three plots may be explained by rater bias. The values for warmth, competence, and valence were determined by averaging the ratings of 10-16 raters (see Data section for more information) for each character. Because all ratings were performed in the 21st century, it is reasonable to assume that the raters had more closely-aligned perceptions of characters commonly used during this time than during previous decades. This may be a factor contributing to the decreasing range of values for these traits over time.

Our next visualization shifts focus slightly away from the perceptions of these characters. Figure 4 shows a dramatic decrease in the total ppm of the top 100 most common characters per decade. In other words, the most popular characters comprise a smaller proportion of all characters being used as time goes on. This would indicate that either some less common characters are becoming more popular, more characters are being used overall, or most likely, both.

Across all of our visualizations, we can see the decrease in usage of the most popular characters and an increased uniqueness of characters in Chinese given names throughout time. The interpretation of patterns seen among the warmth, competence, and valence variables is unclear due to possible bias, but appears to show an increase in the average ppm for characters exhibiting high ratings of these characteristics.

³See Figure 5 in appendix for correlation matrix between per-decade ppm and warmth.

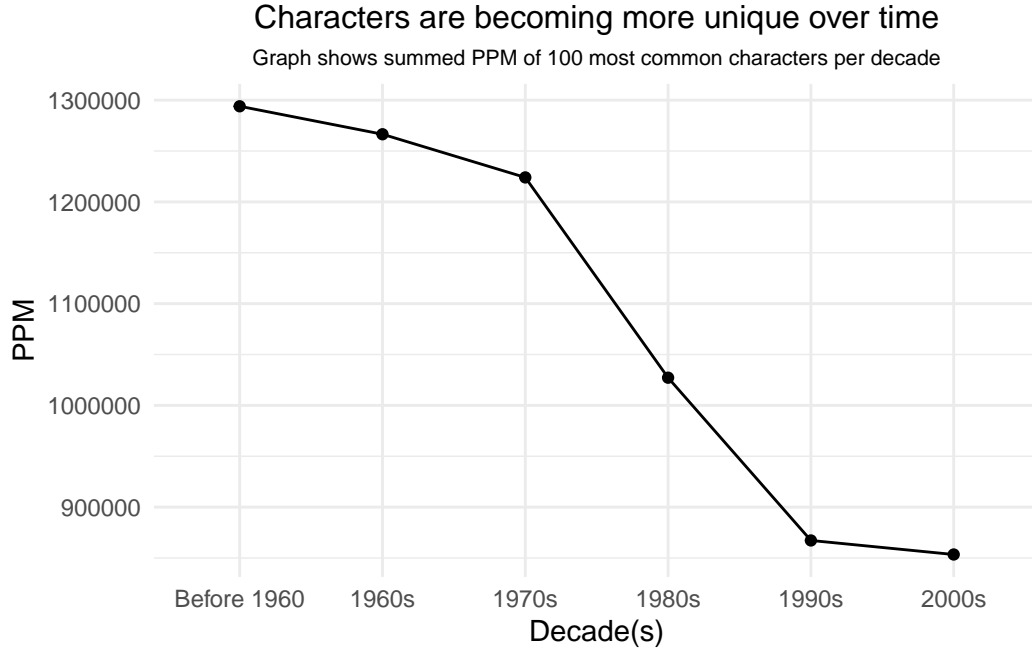


Figure 4: Character Uniqueness Over Time

Discussion

This research project aimed to visualize the change of popularity of Chinese characters in name usage for given names using factors including as ppm, warmth, competence, and valence during the time from 1930-2008. Our visualizations helped reveal an increased uniqueness in characters used in Chinese given names over time. Among all of our visualizations, we can see the decrease in ppm of the most popular characters in Chinese given names throughout time. This could be due to an increase in the popularity of some previously less common characters, an overall increase in the number of characters used in Chinese names, or most likely, both.

The patterns seen among the warmth, competence, and valence variables appear to show an increase in the average ppm for characters exhibiting high ratings of these characteristics. The pattern, however, may also be explained by rater bias. Since these values were determined by averaging rater responses, it is reasonable to presume that they would have more closely-aligned perceptions of characters in current use than during previous decades.

The upward trend in character uniqueness among Chinese names from 1930 to 2008 can be attributed to various underlying factors. Primarily, there has been a cultural shift towards embracing individuality and uniqueness in contemporary society. With globalization and increased diversity, individuals are inclined to set themselves apart by opting for less conventional and more distinct characters for their names. Research done by Xu and Hamamura in 2014

shows that the inclination towards uniqueness is often motivated by the desire to stand out in an increasingly crowded world or to assert personal identity. This trend reflects a departure from the traditional emphasis on fulfilling family wishes in the naming process, signaling a greater emphasis on self-expression and personal choice. This cultural shift reflects a broader societal transformation and an evolving understanding of the role and significance of names in the context of modern life (Xu and Hamamura 2014).

Limitations and Future Research

We had the opportunity to talk with the dataset’s creator and maintainer, Dr. Bruce Bao, who not only answered some of the questions we encountered during our exploratory data analysis but also provided more insight into the field of study. More specifically, regarding the question whether there was a certain set of guidelines followed by the Chinese raters who rated the valence, competence, and the warmth of the names, we learned that each rater gave their own subjective ratings without reference to dictionaries or other outside resources. Therefore, it is reasonable to assume that many factors influence their subjective ratings, such as education level (even a basic understanding of the etymology of most Chinese characters requires a strong academic background), as well as age / birth cohort (different generations may perceive the same Chinese character differently) – leading to both variance and potential bias in the rating process. Nevertheless, these are partially counteracted by including 10+ raters for each attribute and taking the average. Quantitatively, the interrater reliability measures are all above 0.7, which are fairly good. To further reduce the variance in ratings, more raters could be involved, and they could be provided with specific guidelines to follow.

While working with the dataset, we encountered some limitations in the scope of time series data. Ideally, we would have looked at how valence, warmth, and competence changed over time, but the only time series data available dealt with the counts and frequency of characters. This led to a slight redirection in our initial plans, as we were forced to rely on ppm for all longitudinal analysis. While speaking to Dr. Bao, he informed us that the ratings for valence, warmth, and competence are assumed to be stable over time, which allowed us to continue our analysis of these variables. For further research, it would again be useful to increase the number of raters to reduce the variability of these ratings.

Another limitation of our dataset is the fact that it only contains ratings for single Chinese characters. Chinese given names often consist of 2 characters, each with its own meaning, and combining them would yield a different or completely opposite meaning. For example, two characters with negative connotations, when viewed together, would produce a positive connotation. As a result, if we break down the given names into single characters and analyze them individually, we are missing out on some important information pertaining to the original meaning of the name. If we could have the ratings for frequent appearing multi-characters, there might be more interesting findings..

The attempt we made to use a linear regression model to predict name popularity was not very successful and yielded a low R-Squared score, even though we included most of the relevant variables in the dataset as predictors. In addition, the residual plot (difference between predicted name popularity in ppm and actual popularity in ppm vs. actual popularity in ppm) reveals a strong positive linear pattern – the higher the actual popularity, the higher the residual. This indicates the existence of other and potentially more robust variables affecting name popularity that were not included in the dataset we are using. For example, as mentioned in the previous paragraph, ratings for multi-characters might be useful in building a better linear regression model, though how we integrate it with single character ratings would be another issue requiring further examination.

Appendix

For the following plots, blue indicates a positive correlation.

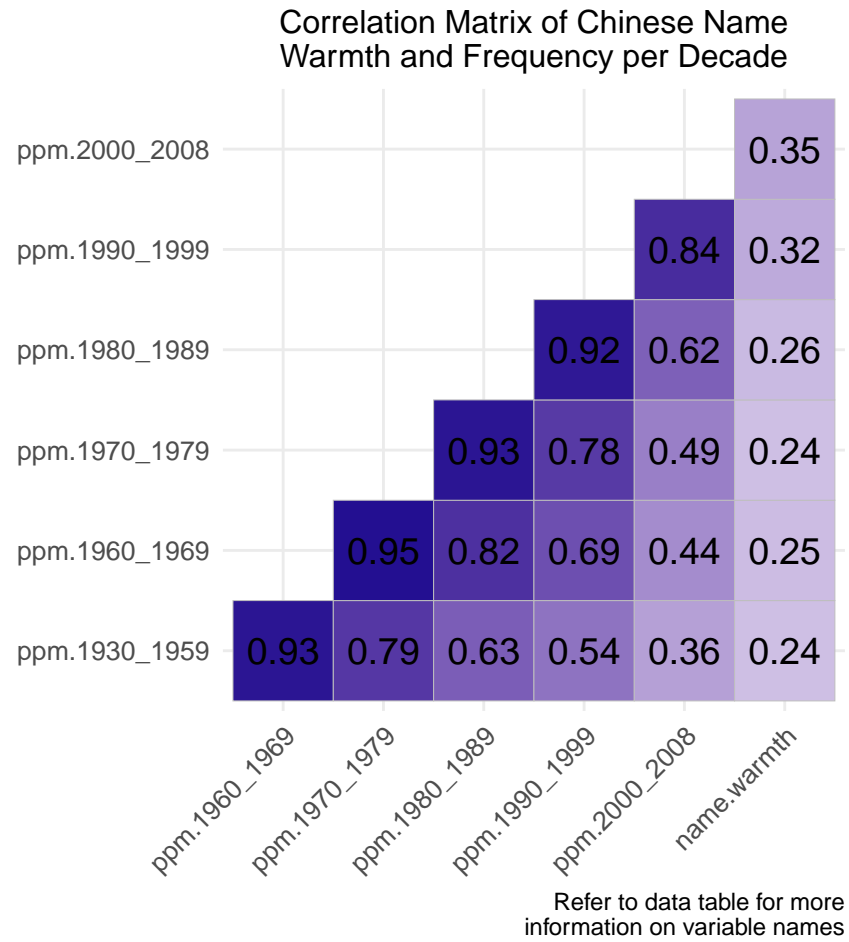
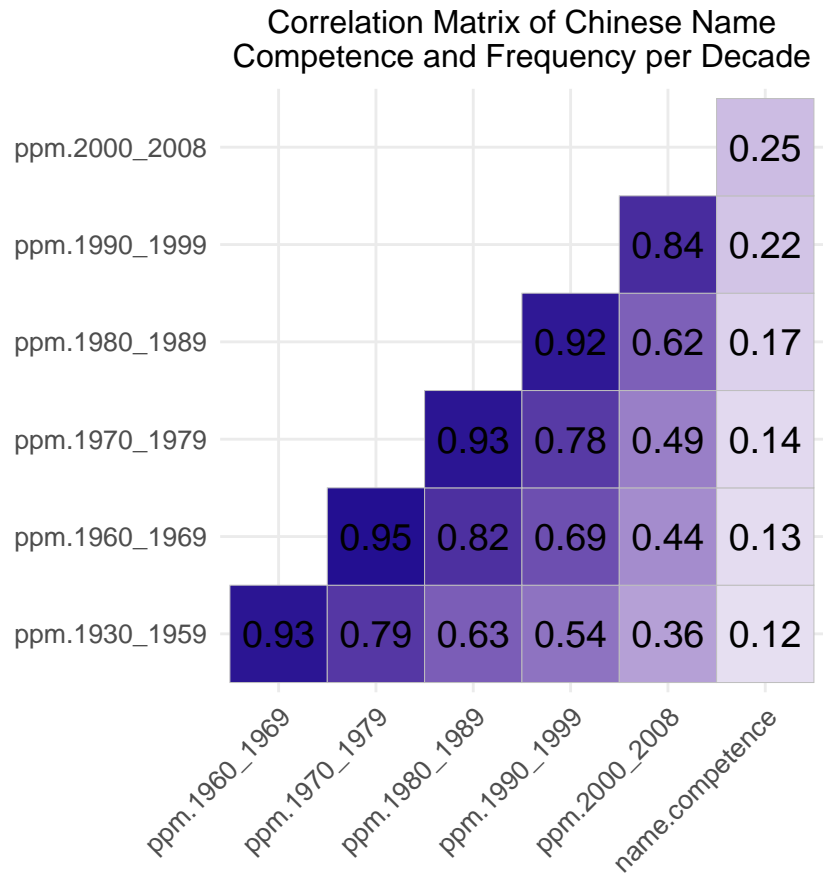


Figure 5: Frequency Correlation Matrix for Warmth



Refer to data table for more
information on variable names

Figure 6: Frequency Correlation Matrix for Competence

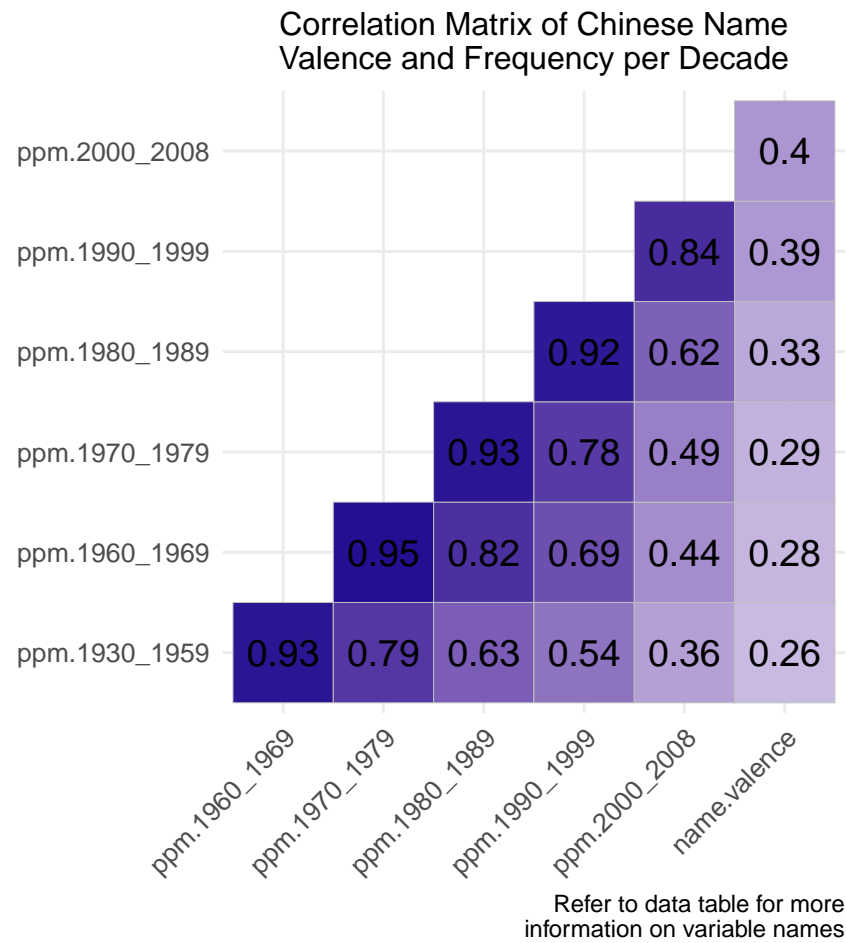


Figure 7: Frequency Correlation Matrix for Valence

References

- Bao, Han-Wu-Shuang. 2021. *ChineseNames: Chinese Name Database 1930-2008*. <https://cran.r-project.org/web/packages/ChineseNames/index.html>.
- Cai, Huajian, Xi Zou, Yi Feng, Yunzhi Liu, and Yiming Jing. 2018. “Increasing Need for Uniqueness in Contemporary China: Empirical Evidence.” *Frontiers in Psychology* 9 (May). <https://doi.org/10.3389/fpsyg.2018.00554>.
- Du, Xiaopeng, Dian Gu, Mengchen Dong, Jiang Jiang, and Wen Jiang. 2021. “Pictographic Name, Warmth Perception, and Trust: Easy Chinese Name Holders Are Seen as Warmer and More Trustworthy.” *Current Psychology* 42 (8): 6846–55. <https://doi.org/10.1007/s12144-021-01938-7>.
- “What Can We Tell from the Evolution of Han Chinese Names?” 2021. <https://kontinentalist.com/stories/a-cultural-history-of-han-chinese-names-for-girls-and-boys-in-china>.
- Wu, Ching-Chao. 1927. “The Chinese Family: Organization, Names, and Kinship Terms.” *American Anthropologist* 29 (3): 316–25. <https://www.jstor.org/stable/661154>.
- Xu, Yi, and Takeshi Hamamura. 2014. “Folk Beliefs of Cultural Changes in China.” *Frontiers in Psychology* 5 (September). <https://doi.org/10.3389/fpsyg.2014.01066>.