University of California, Irvine, Paul Merage School of Business

Master of Science in Business Analytics

- **Class Name:** Statistics for Data Science

- **Report Title:**

  Which country you should visit if you like to kiss?

  - The relationship between GINI index and Kissing Frequency-

- **Submitted by:** Group 6A

- Philip Tong, Pratiksha Gund, Cody Asano, Shing Yang, Xinru Zhao

- **Submitted to:** Professor Rick So

- **Submitted Date:** September 8th, 2024

1. Introduction

We've all experienced love, and human behavior towards love is often far more complex than that of other animals. One of the most notable behaviors is kissing. Humans tend to kiss when they are in love, which is not commonly seen in the behavior of other animals. This raises the question: why do people kiss? It's not just for procreation or the fulfillment of desires but also for forming a deep connection with others, especially in close relationships. In what situations do people tend to kiss more in their relationships? Does the surrounding environment affect kissing behavior? What kinds of relationships encourage more frequent kisses? We explored the relationship between "Kiss Frequency" and various other factors within romantic relationships, such as the attractiveness of a partner, satisfaction with the kiss, and even some economic metrics including national GDP and the GINI index, which will be discussed in the section 3. We believe that there is complexity in this issue because having evidence to prove that at a country level there is a correlation is something that is not really focused on and was something we did not have much knowledge about before we started this analysis project. By examining these factors, we aim to gain a deeper understanding of how both personal and societal influences shape kissing behavior in romantic relationships. These findings will be explored in detail in the subsequent sections of this report.

2. Data Collection

For this research, we sourced data from “Open Science Framework”, which is a collection of pre-written software code, tools, and libraries that are available to the public to use, modify, and distribute (Reference (1)). The data were collected through an online survey with three-thousand one hundred nine participants (Mage = 31.90 years, SD = 11.60 years). Data collection ended after collecting data from 13 countries (6 continents). Participants were recruited from campuses and the wider community, research participant pools, word of mouth, twitter, academic groups on social media and a press release from the lead author’s communications department. No participants were reimbursed for their time. The participants who i) reported being less than 18 years old, ii) did not report their sex as male or female, or, for cross-cultural analyses, iii) if their IP address did not match their reported country of residence were excluded from the data. The final sample size was 2988 participants (794 males, Mage = 32.01 years, SD = 11.56 years).

3. Statistical Analysis & Results

We initially analyzed data from over 2,000 individuals but were unable to identify any significant relationships between the metrics. However, after aggregating the data at the

country level, we could find some correlation between kissing frequencies and one metric, the GINI index variable which represents the income inequality within a country or social group. The data contains across 13 countries and 2,156 sample size (Summarized in Fig.1) . We used the following analytic methods to develop our analysis.

(*Fig.1: Sample size by each country*)

Groups	*Count*
Australia	119
Brazil	170
Chile	141
Colombia	231
CZ	83
France	99
Germany	91
India	43
Italy	60
Nigeria	98
Poland	84
UK	867
USA	70
Total	2156

- 3.1 ANOVA

  Using aggregated data on a national level, we aimed to determine whether there were any differences in kissing frequency across thirteen countries, as outlined in the hypothesis test below:

  - H_0: Kissing frequencies are the same across the 13 countries.
  - H_a: Kissing frequencies are not the same across all 13 countries.

  Since the p-value was smaller than 0.05, we concluded that the mean values of kissing frequency differed across countries, leading us to reject the null hypothesis.

- 3.2 Correlation Analysis

By calculating the correlation coefficient between kissing frequency and various potential influential variables, we found that the highest value, “r=0.64”, was between kissing frequency and the GINI index. The magnitude of r indicated a strong relationship between these two variables. Furthermore, since r>0, it suggests that the GINI index is positively correlated with the frequency of kissing. (Referred in Fig.2.)

(Fig.2 : Correlation coefficient with Kiss Frequencies)

Row Labels	Sex	Age	HPP_9	SRA	SRM	SRH	SRA_Partner	SRM_Partner	SRH_Partner	Kiss_Imp_Initial	Kiss_Imp_Est	Kiss_Freq	Hug_Freq	Sex_Freq	Kiss_Satis	Hug_Satis	Sex_Satis	GK_Breath	GK_Scent	GK_Taste	GK_Wet	GK_Contact	GK_Arous	GK_SynchStyle	GINI	GDP
Australia	1.68	32.47	-0.25	4.37	3.09	4.15	5.46	3.97	4.22	74.66	84.27	78.24	80.51	58.18	76.18	82.44	64.46	83.88	77.18	71.43	65.59	73.75	66.94	61.63	30.30	49900
Brazil	1.67	29.13	0.93	4.12	3.42	4.35	4.91	3.98	4.46	83.78	86.46	80.26	77.68	66.72	79.36	84.58	69.35	90.52	88.57	80.04	67.66	85.90	80.77	78.14	49.70	15500
Chile	1.82	31.15	-0.45	4.61	3.02	4.42	4.97	4.50	4.50	84.12	85.29	81.58	85.94	66.40	80.40	79.83	73.66	89.13	87.82	81.12	66.28	80.56	78.20	65.76	50.50	24600
Colombia	1.67	26.15	0.27	4.45	3.25	4.38	5.12	4.02	4.63	78.45	83.55	79.20	82.47	68.87	81.31	81.00	76.31	92.71	89.68	81.66	67.84	74.84	78.15	71.61	53.50	14500
CZ	1.81	27.24	-0.87	4.73	3.11	4.47	5.36	4.45	4.93	82.27	78.59	75.05	80.08	62.25	77.37	78.10	69.57	83.55	81.20	72.84	65.64	78.99	73.83	64.38	25.00	35200
France	1.77	32.25	-0.46	4.72	3.91	4.59	5.29	4.01	4.77	85.00	75.39	70.46	76.71	65.21	74.18	78.91	72.53	87.41	82.38	74.70	55.78	73.75	80.04	55.86	29.20	43600
Germany	1.85	27.29	-0.87	4.46	2.85	4.47	5.27	4.27	4.63	84.90	82.75	74.63	82.30	60.71	83.24	81.51	72.48	82.67	81.98	73.04	66.24	70.45	60.19	69.10	27.00	50200
India	1.86	28.67	0.94	4.79	2.90	4.12	5.24	4.76	4.90	75.79	82.84	72.30	75.07	52.58	77.70	73.69	62.92	85.44	83.26	74.57	65.32	85.72	85.07	61.85	35.20	7200
Italy	1.87	30.23	0.16	4.05	3.39	4.68	5.13	4.55	4.69	85.40	85.90	74.40	66.56	68.85	77.47	75.46	74.56	86.32	86.24	83.28	63.93	82.35	83.27	70.53	31.90	38000
Nigeria	1.62	31.92	1.16	5.52	3.01	6.01	5.74	4.62	5.94	58.11	79.76	78.33	79.01	53.18	82.66	80.20	71.03	86.00	83.00	80.05	65.12	83.10	85.27	60.84	48.80	5900
Poland	1.74	27.63	-0.87	4.72	3.30	4.73	5.48	4.54	4.87	82.63	85.11	77.29	88.87	67.65	79.00	82.54	71.54	85.43	82.71	77.27	62.19	71.85	77.48	64.30	30.80	29300
UK	1.74	36.65	-1.06	4.22	3.01	4.29	5.43	4.21	4.68	84.20	86.33	80.76	79.28	64.95	78.47	82.61	71.88	86.54	81.19	74.53	65.49	75.93	70.99	63.82	32.40	43600
USA	1.77	32.76	-0.89	4.80	3.12	4.41	5.39	4.45	4.80	78.79	85.77	82.61	77.33	66.25	82.89	86.23	73.63	84.83	82.09	77.09	62.77	73.59	71.03	68.30	45.00	59500
Grand Total	1.74	32.36	-0.41	4.43	3.13	4.44	5.32	4.25	4.71	81.45	84.46	78.98	79.91	64.44	79.12	81.64	71.74	87.15	83.33	76.56	65.28	77.04	74.39	65.87	38.04	33744
Correlation Coefficient with Kiss_Freq	-0.44	0.31	-0.11	-0.08	-0.33	0.00	-0.06	-0.12	-0.13	-0.12	0.66	1.00	0.36	0.30	0.55	0.72	0.27	0.31	0.21	0.32	0.50	-0.02	-0.19	0.45	0.64	0.04

- 3.3 Simple Linear Regression

We conducted simple linear regression as outlined in the hypothesis test below:

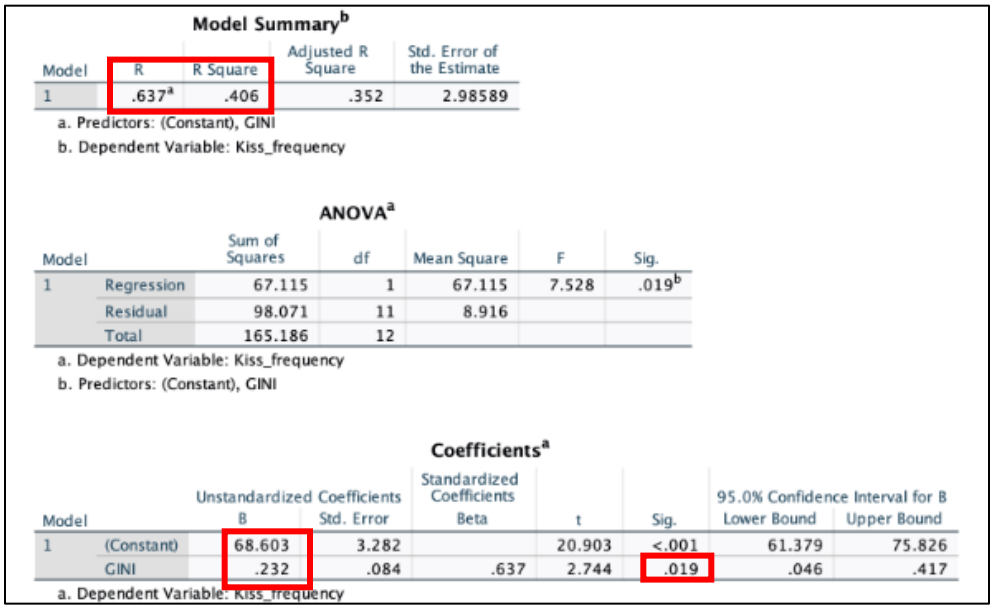
- H_0: Kissing frequencies are the same across the 13 countries.
- H_a: Kissing frequency does not decrease as GINI index increases

The regression model is as below.

$$\underline{\text{KissingFreq} = 68.603 + 0.232 * \text{GINI}}$$

The R squared value was 0.406, and the p-value of 0.019 was smaller than 0.05, indicating that approximately 40.6% of the total variation could be explained by the regression line, and the GINI coefficient was statistically significantly different from zero. This provided strong evidence of a linear relationship between kissing frequency and the GINI index. The model effectively analyzed their relationship, showing that for each one-unit increase in GINI, kissing frequency would increase by 0.232. In other words, as the GINI coefficient rises, kissing frequency tends to increase. (Summarized in Fig.3).

(*Fig.3: Simple linear regression model summary*)



Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.637[a]	.406	.352	2.98589

a. Predictors: (Constant), GINI
b. Dependent Variable: Kiss_frequency

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	67.115	1	67.115	7.528	.019[b]
	Residual	98.071	11	8.916		
	Total	165.186	12			

a. Dependent Variable: Kiss_frequency
b. Predictors: (Constant), GINI

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	68.603	3.282		20.903	<.001	61.379	75.826
	GINI	.232	.084	.637	2.744	.019	.046	.417

a. Dependent Variable: Kiss_frequency

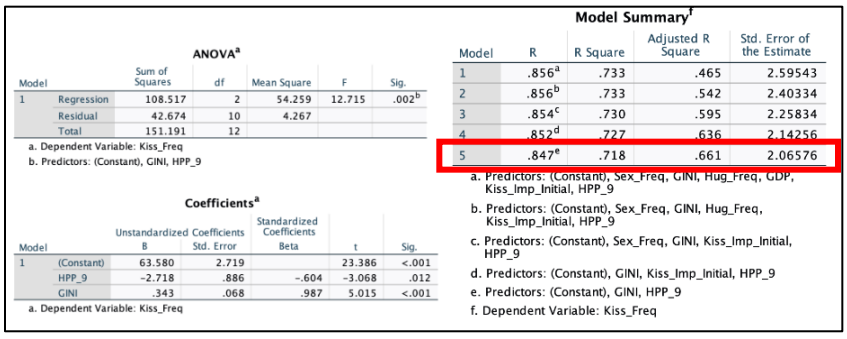
- 3.4 Multiple Regression

Although the initial conclusions were solid, we sought to further improve our model. To do so, we added variables such as Sex Frequency, Hug Frequency, GDP, Kiss Impression Initial, and HPP_9 (Historical Pathogen Prevalence, which refers to the immunity of a country to disease). This variable may have an impact on close interaction between people, and kissing is one of these interactions. Using multiple regression, we derived the following equation:

$$\underline{\text{KissingFreq} = 63.58 + 0.343 \times \text{GINI} - 2.718 \times \text{HPP\_9}}$$

After applying the backward method for variable selection, we achieved an adjusted R squared of 0.661 and a p-value of 0.002. This indicated that the model performed well and was statistically significant, with GINI and HPP_9 as independent variables. The p-values for both GINI and HPP_9 were less than 0.05, confirming a positive linear relationship between kissing frequency and GINI, and a negative linear relationship between kissing frequency and HPP_9. (Summarized in Fig.4).

(*Fig.4: Simple linear regression model summary*)



ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	108.517	2	54.259	12.715	.002[b]
	Residual	42.674	10	4.267		
	Total	151.191	12			

a. Dependent Variable: Kiss_Freq
b. Predictors: (Constant), GINI, HPP_9

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	63.580	2.719		23.386	<.001
	HPP_9	-2.718	.886	-.604	-3.068	.012
	GINI	.343	.068	.987	5.015	<.001

a. Dependent Variable: Kiss_Freq

Model Summary[f]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.856[a]	.733	.465	2.59543
2	.856[b]	.733	.542	2.40334
3	.854[c]	.730	.595	2.25834
4	.852[d]	.727	.636	2.14256
5	.847[e]	.718	.661	2.06576

a. Predictors: (Constant), Sex_Freq, GINI, Hug_Freq, GDP, Kiss_Imp_Initial, HPP_9
b. Predictors: (Constant), Sex_Freq, GINI, Hug_Freq, Kiss_Imp_Initial, HPP_9
c. Predictors: (Constant), Sex_Freq, GINI, Kiss_Imp_Initial, HPP_9
d. Predictors: (Constant), GINI, Kiss_Imp_Initial, HPP_9
e. Predictors: (Constant), GINI, HPP_9
f. Dependent Variable: Kiss_Freq

4. Discussion & Conclusion

In conclusion, the stresses of income inequality often drive individuals to seek comfort and emotional support from their partners. In this context, kisses become more than mere gestures of affection—they serve as vital expressions of emotion and a source of solace. As financial disparities create heightened emotional strain, couples may turn to intimate acts like kissing to reaffirm their bond and find relief from their challenges. This behavior highlights how deeply personal and societal stresses are intertwined, with kisses acting as tangible means of coping and connection amidst broader economic pressures. Notably, kissing frequency varies across different countries, which may reflect varying levels of economic and emotional stress. To gain further insight, correlating kissing frequency with health metrics such as HPP_9—an index reflecting a country's immunity to disease and overall health—can provide a more comprehensive understanding of how economic pressures and health impact intimate behaviors.

5. Reference

(1) Open Science Framework: https://osf.io/pbqwm/