



CORRESPONDENCE ANALYSIS

Young People Survey

GAGANA UDAY KUMAR

WOV797



OBJECTIVE

ANALYZING THE RELATIONSHIP BETWEEN
EDUCATION AMONG THE YOUNG PEOPLE AND
THEIR INTERNET USAGE

DATA SOURCE

- The data file is a survey data which consists of 1010 rows and 150 columns.
- All participants were aged between 15-30
- The variables can be split into the following groups: Music preferences, Movie preferences, Hobbies & interests, Phobias, Health habits, Personality traits, views on life, & opinions, Spending habits, Demographics

A	B	C	
ID	Education	Internet_Usage	
1	4	3	
2	4	3	
3	3	3	
4	4	4	
5	3	3	
6	3	3	
7	3	2	
8	4	3	
9	3	3	
10	3	3	
11	3	2	
12	2	3	
13	4	3	

EDUCATION VARIABLE

currently a primary school pupil

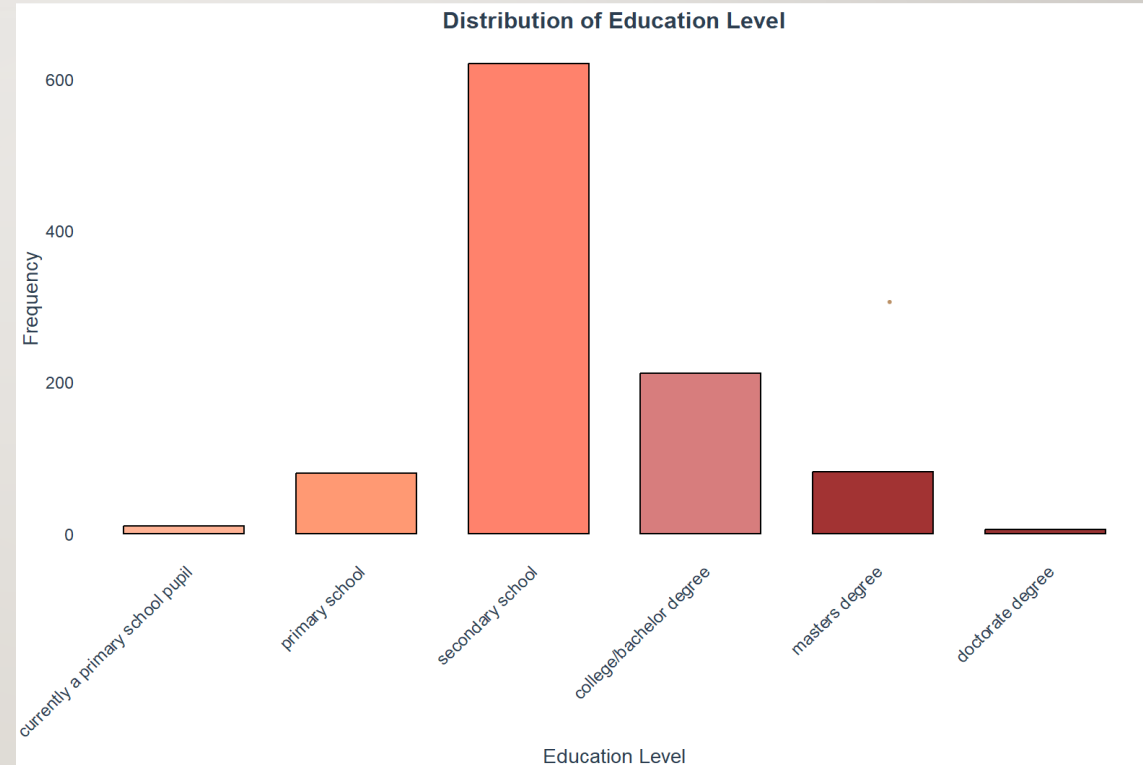
primary school

secondary school

college/bachelor degree

Master's degree

doctorate degree



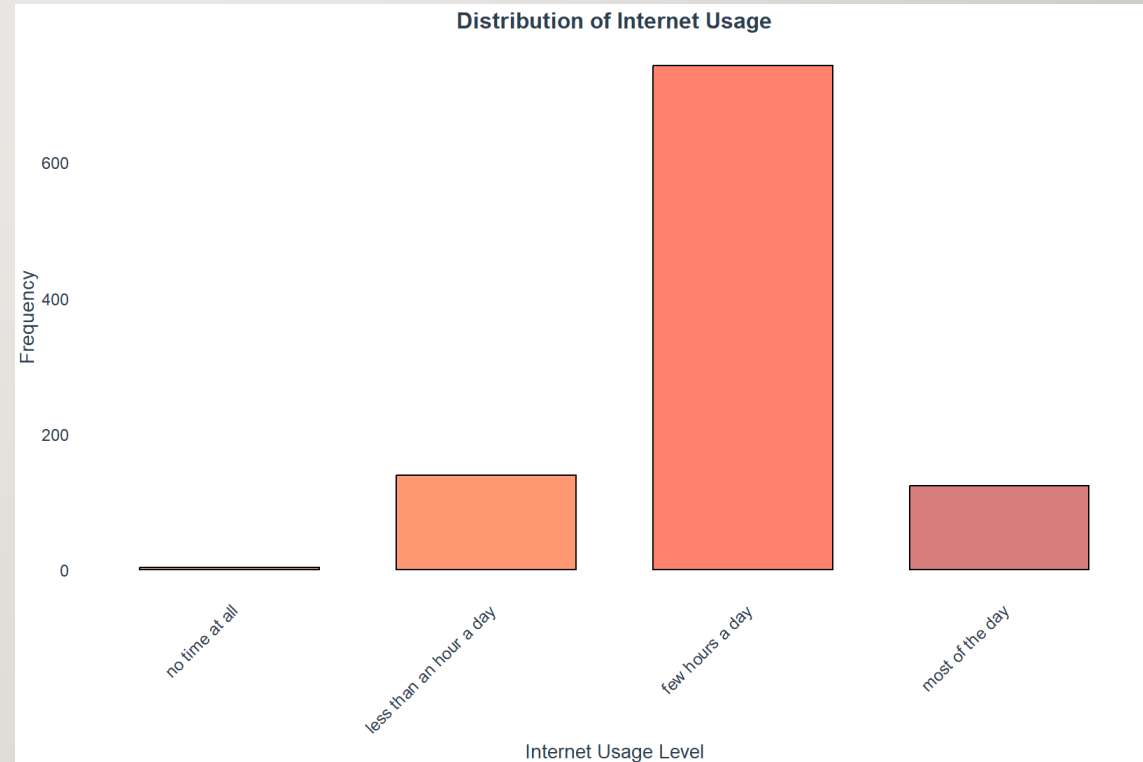
INTERNET USAGE VARIABLE

no time at all

less than an hour a day

few hours a day

most of the day



WHAT IS CORRESPONDENCE ANALYSIS

- Correspondence Analysis (CA) is a multivariate statistical technique used for analyzing the relationships between categorical variables in a contingency table.
- It is particularly useful for exploring the associations between two or more categorical variables and visualizing these relationships in a low-dimensional space.

CONTINGENCY TABLE

Contingency Table					
	few_hours_a_day	less_than_an_hour_a_day	most_of_the_day	no_time_at_all	Sum
college_or_bachelor_degree	155	25	32	0	212
currently_a_primary_school_pupil	5	4	1	0	10
doctorate_degree	4	1	0	0	5
masters_degree	51	17	11	2	81
primary_school	62	8	10	0	80
secondary_school	466	84	70	1	621
Sum	743	139	124	3	1009

EXPECTED VS OBSERVED - EXPECTED

Chi-Square Statistic Expected Values				
	few_hours_a_day	less_than_an_hour_a_day	most_of_the_day	no_time_at_all
college_or_bachelor_degree	156.111	29.205	26.054	0.630
currently_a_primary_school_pupil	7.364	1.378	1.229	0.030
doctorate_degree	3.682	0.689	0.614	0.015
masters_degree	59.646	11.159	9.954	0.241
primary_school	58.910	11.021	9.832	0.238
secondary_school	457.287	85.549	76.317	1.846

Observed Minus Expected Values				
	few_hours_a_day	less_than_an_hour_a_day	most_of_the_day	no_time_at_all
college_or_bachelor_degree	-1.11100	-4.20515	5.94648	-0.63033
currently_a_primary_school_pupil	-2.36373	2.62240	-0.22894	-0.02973
doctorate_degree	0.31814	0.31120	-0.61447	-0.01487
masters_degree	-8.64618	5.84143	1.04559	1.75917
primary_school	3.09019	-3.02081	0.16848	-0.23786
secondary_school	8.71259	-1.54906	-6.31715	-0.84638

CHI-SQUARE CONTRIBUTION

Contributions to the Total Chi-Square Statistic					
	few_hours_a_day	less_than_an_hour_a_day	most_of_the_day	no_time_at_all	Sum
college_or_bachelor_degree	0.0079	0.6055	1.3572	0.6303	2.6010
currently_a_primary_school_pupil	0.7587	4.9920	0.0426	0.0297	5.8231
doctorate_degree	0.0275	0.1406	0.6145	0.0149	0.7974
masters_degree	1.2533	3.0579	0.1098	12.8499	17.2710
primary_school	0.1621	0.8280	0.0029	0.2379	1.2309
secondary_school	0.1660	0.0280	0.5229	0.3880	1.1049
Sum	2.3756	9.6521	2.6500	14.1507	28.8283

The SAS System

The CORRESP Procedure

Inertia and Chi-Square Decomposition

Singular Value	Principal Inertia	Chi-Square	Percent	Cumulative Percent	0	20	40	60
0.14473	0.02095	21.1357	73.32	73.32				
0.07215	0.00521	5.2523	18.22	91.54				
0.04918	0.00242	2.4403	8.46	100.00				
	0.02857	28.8283	100.00					

Pr > Chisq = .0169 DF = 15

CHI-SQUARE

SUMMARY STATISTICS FOR ROW POINTS

Summary Statistics for the Row Points			
	Quality	Mass	Inertia
college_or_bachelor_degree	0.5236	0.2101	0.0902
currently_a_primary_school_pupil	0.9241	0.0099	0.2020
doctorate_degree	0.5332	0.0050	0.0277
masters_degree	0.9996	0.0803	0.5991
primary_school	0.9848	0.0793	0.0427
secondary_school	0.6733	0.6155	0.0383

SUMMARY STATISTICS FOR COLUMN POINTS

Summary Statistics for the Column Points			
	Quality	Mass	Inertia
few_hours_a_day	0.8324	0.7364	0.0824
less_than_an_hour_a_day	0.9906	0.1378	0.3348
most_of_the_day	0.3582	0.1229	0.0919
no_time_at_all	0.9823	0.0030	0.4909

BONUS - SQUARED COSINES

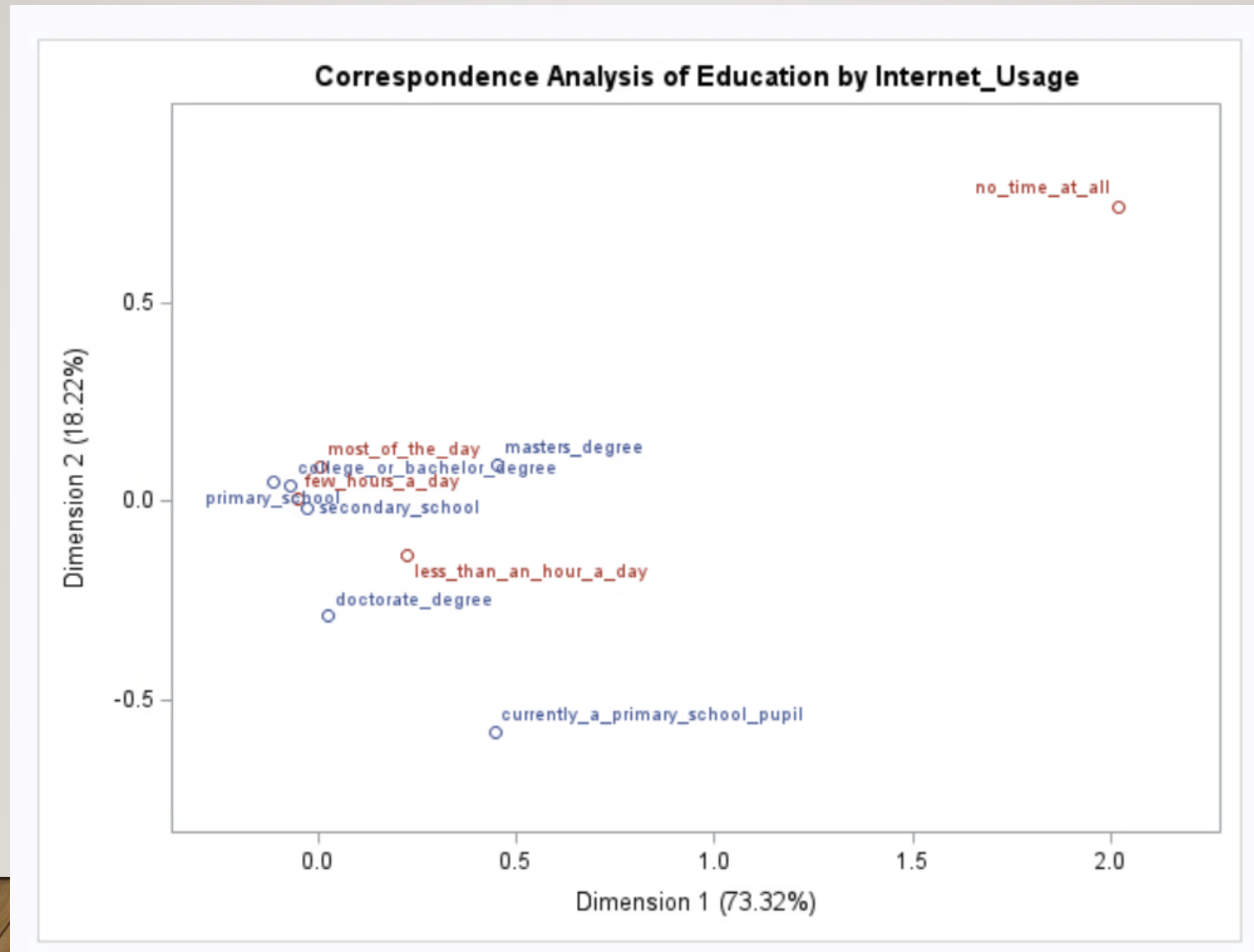
Squared Cosines for the Row Points

	Dim1	Dim2
college_or_bachelor_degree	0.3905	0.1330
currently_a_primary_school_pupil	0.3418	0.5822
doctorate_degree	0.0047	0.5285
masters_degree	0.9613	0.0383
primary_school	0.8356	0.1492
secondary_school	0.4476	0.2257

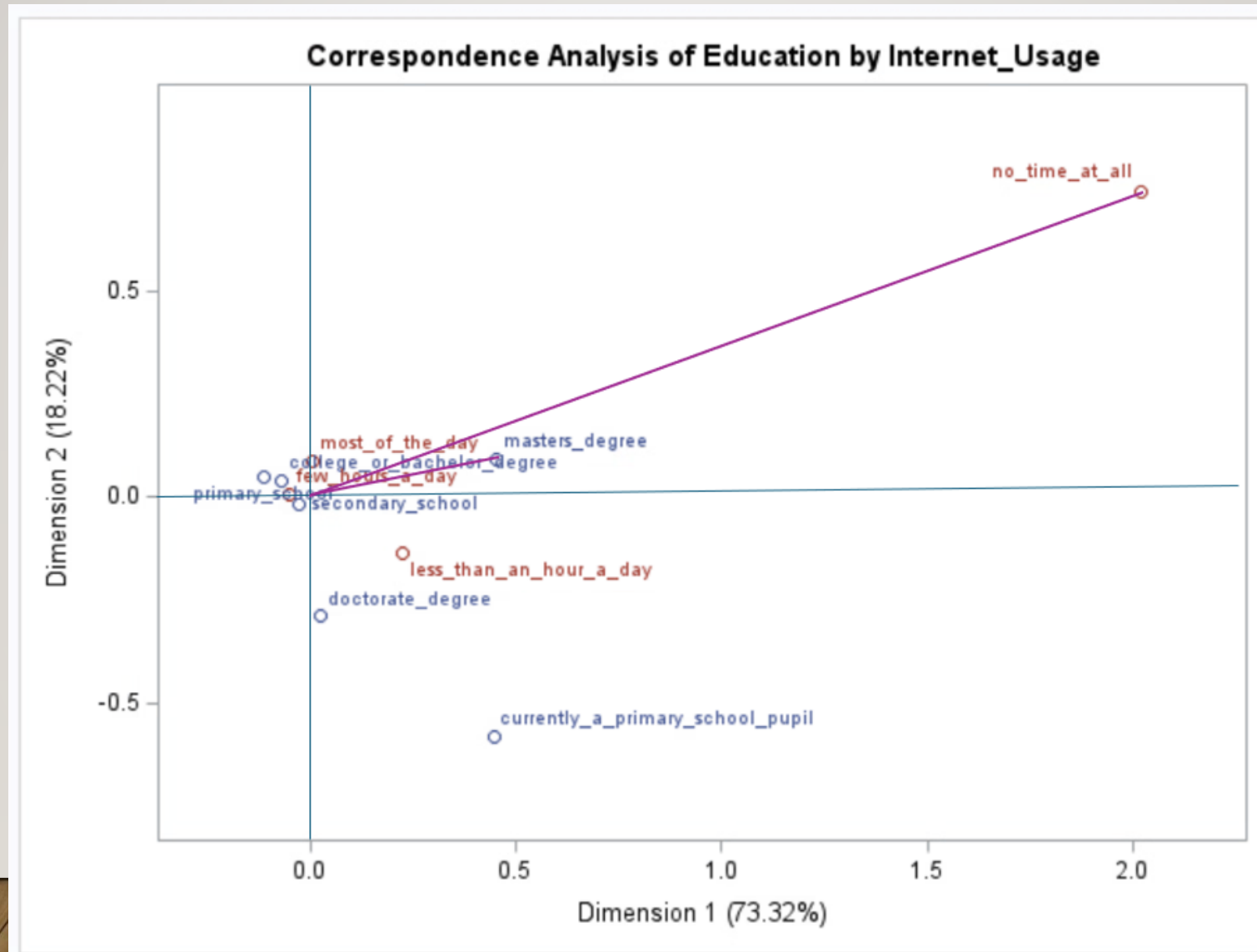
Squared Cosines for the Column Points

	Dim1	Dim2
few_hours_a_day	0.8121	0.0202
less_than_an_hour_a_day	0.7207	0.2699
most_of_the_day	0.0015	0.3567
no_time_at_all	0.8654	0.1169

CORRESPONDENCE MAP



EXAMPLE OF STRONG ASSOCIATION



CONCLUSION

- **Association between Education and Internet Usage:** The analysis reveals a relationship between education levels and internet usage patterns.
- **Specific Trends:** Higher education levels, such as master's degrees, tend to be associated with little to no internet usage.
- **Influence of Education:** Education significantly shapes internet usage behavior, indicating the need for targeted interventions.
- **Variability in Usage:** There is notable variation in internet usage across education levels, affecting access to information and opportunities.
- **Policy Implications:** The findings can guide policies to improve digital access and literacy, bridging gaps across different education levels.