BACHELOR'S DEGREE THESIS

Degree in Statistics

Title: The digital practices of the elderly population, an international comparison

Author: Lu Li

Advisor: Mireia Fernández Ardèvol

Department: Dept. Econometrics, Statistics and Applied Economics

Academic year: 2017-2018

UNIVERSITAT DE BARCELONA

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

ABSTRACT
The popularity of the Internet has changed the way people live. However, we know very little about the Internet use of older people. This study focuses on senior's Internet and mobile phone use in three countries. Data come from an online survey in Canada (3538), Spain (2238) and the Netherlands (801). The latent class analysis is used to clustering the older internet users in latent (unobservable) classes through the analysis of their patterns in categorical observed indicators. This research indicates individuals with higher income and higher educational level show higher use of digital tools, either the internet o the mobile phone, in their everyday life.

KEYWORDS: Latent class analysis (LCA); Online survey; Older people; Internet use; Mobile phone use; International comparison.

RESUMEN
La popularidad de Internet ha cambiado la forma de vivir de las personas. Sin embargo, sabemos muy poco sobre el uso de Internet de las personas mayores. Este estudio se centra en el uso de Internet y teléfonos móviles de personas mayores en tres países. Informa los resultados de una encuesta en línea para el proyecto "Estudio longitudinal transnacional: audiencias mayores en el entorno de medios digitales" realizado en Canadá (3538 individuos), España (2238) y los Países Bajos (801) en 2016. Dirigido a usuarios mayores de Internet que tienen 60 años o más, sin límite superior de edad. El análisis de clase latente se utiliza para agrupar a las personas mayores usuarios de internet en clases latentes (no observables) mediante el análisis de sus patrones en indicadores categóricos observados. Esta investigación indica que hay diversidad de usos digitales entre personas mayores. En los tres países, se confirma que las características socioeconómicas influyen a los usos digitales. Como se esperaba, las personas con mayores ingresos y nivel educativo superior muestran un mayor uso de herramientas digitales en su vida cotidiana. Además, el uso de Internet está positivamente asociado con el uso del teléfono móvil; Las personas mayores que utilizan los medios e Internet a menudo muestran un uso más intenso del teléfono móvil.

PALABRAS CLAVE: Análisis de clase latente (LCA); Encuesta en línea; Personas mayores; Uso de Internet; Uso del teléfono móvil; Comparación internacional.

AMS CLASIFICATON
62P25- Applications to social sciences

Index

I. INTRODUCTION

This study reports the results of an online survey for the project "Cross-national longitudinal study: Older audiences in the digital media environment" conducted in Canada, Spain and the Netherlands in 2016 (the countries are ordered by their sample size). The universe of study of these surveys is the online population aged 60 and over, with no upper threshold age. Is why an online survey is appropriated tool for data gathering. In addition, the tool is appropriate to analyse digital everyday life practices around media, internet and mobile phone use.

The main goal of the study is characterize the digital practices of the Internet users of 60 years and over in Canada, Spain and the Netherlands. Moreover, solve the research questions that lead the analysis: Can we find differentiated socio-demographic characteristics among elderly people in relation to the use of digital media and the internet, and in relation to the use of the mobile phone? Is there a common tendency in Canada, Spain and the Netherlands?

The use of internet among people aged 65 years or more has grown by 150% in the last six years (2010-2016), 35% of people between 65 and 74 are Internet users, according to Eurostat data. (UOC News, 2017). In addition, according the Statistics Canada, the survey titled Canadians at Work and Home found online activity among those aged 65 to 74 climbed 16 percentage points between 2013 and 2016. That growth was closely followed among people aged 75 and older, whose internet use jumped 15 percentage points over the same three-year period (CTV News, 2017). Due to the fast growing trend of adoption, it is relevant to analyses the digital practices of older population.

This study consists in 3 main parts, first it includes the methodology that explains the data reference, database structure, statistical techniques for the data treatment and data analysis. The latent class analysis is the main statistical technique applied to data analysis that was used to clustering the individuals in latent (unobservable) classes through the analysis of their patterns in categorical observed indicator. The second part includes comparison results of univariate description analysis and latent class analysis of the three countries. Thirdly, the conclusion section discusses the entire results of the study. In addition, the appendix contains tables, charts and the results of hypothesis contrasts obtained from RStudio.

Finally, I want to thank my family and my friends for their support. I would like to thank in particular my tutor Mireia Fernández Ardèvol for her unconditional support, her valuable advice and taking time from her busy schedule to help me with this final degree project. I also want to thank acknowledge the Ageing + Coummunication +Tecnologies (ACT) project for sharing the database I analyze (http://actproject.ca/, SSHRC Canada ref.: 895-2013-1018).

II. METHOD

II.1. Origin of the data

The database comes from the project "Cross-National Longitudinal Study: "Older Audiences in the Digital Media Environment", an online survey targeted to elder Internet users (60 years old and over). Three countries were selected for this study: Canada (3538 responses), Spain (2238) and the Netherlands (801). Data were collected between November and December 2016, with the exception of the Canadian data that were collected between June and July 2017 (Loss, Nimrod & Fernández-Ardèvol, 2018).

II.2 Data reference

The rows (individuals) of the database refer to each respondent. The columns (variables or attributes) can be divided into three blocks: the first block gathers the socio demographic variables, such as age, gender, level of education; the second refers to the use of media and internet and the third one to the use of mobile phone. In **Appendix Table A.1**, reproduces the original survey questions selected for this study. In addition, **Appendix Table A.2**, contains a description of all the variables corresponding to selected questions.

II.3 Statistical techniques for the treatment of data

II.3.1 Recoding and regrouping of the variables

Regarding the socio-demographic variables, the ages are grouped into 5 categories with a range of 5 years in each interval. For education, I have built a new variable with 4 categories that are: "Primary or less" (up to 8-9 years of education), "Secondary" (between 10 and 14 years of education), "Tertiary" (15 years of education or more) and "Don't know". Relating to the employment, the variable has 5 categories where the category "Active" represents full-time work and part-time work; "Inactive" represents retired, unemployed or in unpaid position (household, volunteer or community service). For family status, it is divided into two variables "has married" and "has children" with 3 categories "Yes", "No" and "NR" where "NR" means preferred not to respond. Also for income the new constructed variable contains 4 categories "Above the average", "Similar to average", "Below the average" and "Not declared".

The time spent for each media in this study is not important; I will mainly focus on seniors' Internet use behaviour patterns. So I deleted the category "hours and minutes" from all the variables of the second block (Media &Internet Usage).

In the blocks of media & internet usage, the structure of the survey answers is negative logic, for example, the question is: "Please thinking of yesterday; how much time did you spend on the following media?" The options of the answers are: hours and minutes, didn't use (Yes or No) and don't remember (Yes or No). I decided to recode the categories "didn't use" and "don't remember", because if the respondent answers "Yes" in "Didn't use?" it means he/she has not used it, but I want to use "Yes" to express that he/she used it to avoid possible confusion. Besides, I merged the category "didn't use" and “don´t remember” into one since it has repeated information and converted them into dichotomous variables, with the names of the categories “Yes” and “Other”, which "Yes" represents those who used and remembered, “Other” means didn’t use but remember or used but can't remember.

Finally, all variables related to mobile phone usage will remain the same; I just added labels for them.

Table II.1 contains detailed information of variables that are considered necessary for the analysis.

Table II. 1 Description of variables that are considered necessary for the analysis

	Variable	code	Label	Meaning
SOCIO-DEMOGRAPHIC VARIABLES	Q21_Sex	1	Male	Gender
		2	Female	
	Q22_Age	Numeric variable		Age
	Q22_Age_5cat	1	[60, 65[	Divide age into 5 categories
		2	[65, 70[	
		3	[70, 75[	
		4	[75, 80[	
		5	[80, +]	
	has_partner	1	Married	Are you married?
		2	Not Married	
		3	NR	
	Has_children	1	Yes	Do you have children?
		2	No	
		3	NR	
	edu_3cat	1	Primary or less	The education status
		2	Secondary	
		3	Tertiary	
		4	Don't know	
	income_3cat	1	Above the average	Income categories
		2	Similar to average	
		3	Below the average	
		4	Not declared	
	employ_3cat	1	Active	Employment status
		2	Inactive	
		3	Other	
		4	DK_NR	
	tvset	1	Yes	Watched television on a tv set
		0	Other	
	tvcom	1	Yes	Watched television on a computer
		0	Other	
	tvmob	1	Yes	Watched television on a mobile phone
		0	Other	
	radset	1	Yes	Listened to radio on a radio set
		0	Other	
	radmob	1	Yes	Listened to radio on a mobile phone
		0	Other	
	radcomp	1	Yes	Listened to radio on a computer
		0	Other	
	newsint	1	Yes	Read newspapers or magazines on the internet
		0	Other	

MEDIA & INTERNET USAGE	newsprint	1	Yes	Read newspapers or magazines in the printed version
		0	Other	
	booksprint	1	Yes	Read books in the printed version
		0	Other	
	bookselec	1	Yes	Read books in the electronic version
		0	Other	
	audbooks	1	Yes	Listened to audio books
		0	Other	
	internews	1	Yes	Internet use yesterday / Getting news
		0	Other	
	interemails	1	Yes	Internet use yesterday / writing and reading e-mails
		0	Other	
	interpodcast	1	Yes	Internet use yesterday / Downloading music, film or podcasts
		0	Other	
	intergames	1	Yes	Internet use yesterday / Playing computer games online
		0	Other	
	interSNS	1	Yes	Internet use yesterday / Using social network sites
		0	Other	
	interchat	1	Yes	Internet use yesterday / Using chat programs
		0	Other	
	interreadblogs	1	Yes	Internet use yesterday / Reading entries at debate sites, blogs
		0	Other	
	intershopping	1	Yes	Internet use yesterday / Online shopping, banking, travel reservation etc.
		0	Other	
	interhobbies	1	Yes	Internet use yesterday / Using websites concerning my interests or hobbies
		0	Other	
	interother	1	Yes	Internet use yesterday / Other
		0	Other	
MOBILE PHONE USAGE	sms	1	Yes	SMS
		0	No	
	mms	1	Yes	Multimedia Message Services
		0	No	
	radio	1	Yes	Listening to radio
		0	No	
	podcast	1	Yes	Listening to podcast
		0	No	
	music_player	1	Yes	Using a phone as a music player
		0	No	
	photos	1	Yes	Taking photographs
		0	No	
	rec_video	1	Yes	Recording video
		0	No	
	web_browser	1	Yes	Viewing websites via browser
		0	No	
	web_apps	1	Yes	Viewing websites via apps
		0	No	
	inst_mess	1	Yes	Instant messaging
		0	No	
	sns	1	Yes	Social network sites
		0	No	
	games	1	Yes	Games
		0	No	

	calendar	1	Yes	Calendar
		0	No	
	alarm	1	Yes	Alarm clock and reminders
		0	No	
	email	1	Yes	E-mail
		0	No	
	gps_maps	1	Yes	GPS and maps
		0	No	
	down_apps	1	Yes	Downloading apps
		0	No	
	voice_calls	1	Yes	Ordinary voice call
		0	No	
	other	1	Yes	Other
		0	No	
OTHERS	ID	Numeric variable		ID merged dataset
	Countrycode	2	Canada	Country code
		5	Netherlands	
		7	Spain	
	POND	Numeric variable		Ponderation for Canada

II.3.2 Structure of the data matrix

The initial database matrix consists of:
- Individuals: 6577
- Variables: 117

After regrouping and remove variables not relevant for this research, the final data matrix has:
- Individuals: 6577
- Variables: 52
  - Qualitative variable: 49
  - Quantitative variables: 3 (ID, Q22_Age, POND)

In the database, there are missing values in all the variables of mobile usage because there are respondents who do not have a mobile. Therefore, the number of missing in each variable in this block (mobile phone usage) is equal to 864, which distributed in three countries.

Table II. 2 Missing values in the mobile usage block by country

Country	Missing in the block of mobile usage	Total sample	% of missing in each country
Canada	620	3,538	17.52%
Spain	215	2,238	9.61%
Netherlands	29	801	3.62%
Total	864	6577	100%

II.3.3 Outliers treatment

- Z- Score method

As Pulletikurti said (Pulletikurti, 2015), observed variables often contain outliers that have unusually large or small values when compared with others in a data set. Some data sets may come from homogeneous groups; others from heterogeneous groups that have different characteristics regarding a specific variable. Outliers can be caused by incorrect measurements, including data entry errors, or by coming from a different population than the rest of the data. The deleterious effects of outliers on statistical analyses are:

1) Outliers generally serve to increase error variance and reduce the power of statistical tests.

2) If non-randomly distributed, they can decrease normality, altering the odds of making both Type I and Type II errors.

3) They can seriously bias or influence estimates that may be of substantive interest.

Before processing the data I have divided the database between 3 countries (Canada, Spain and the Netherlands) using the split () function of R. Furthermore, outliers have been detected in the age variable since it is the unique quantitative variable, as shown in the followings charts:

Figure II.1 Age distribution of Canada

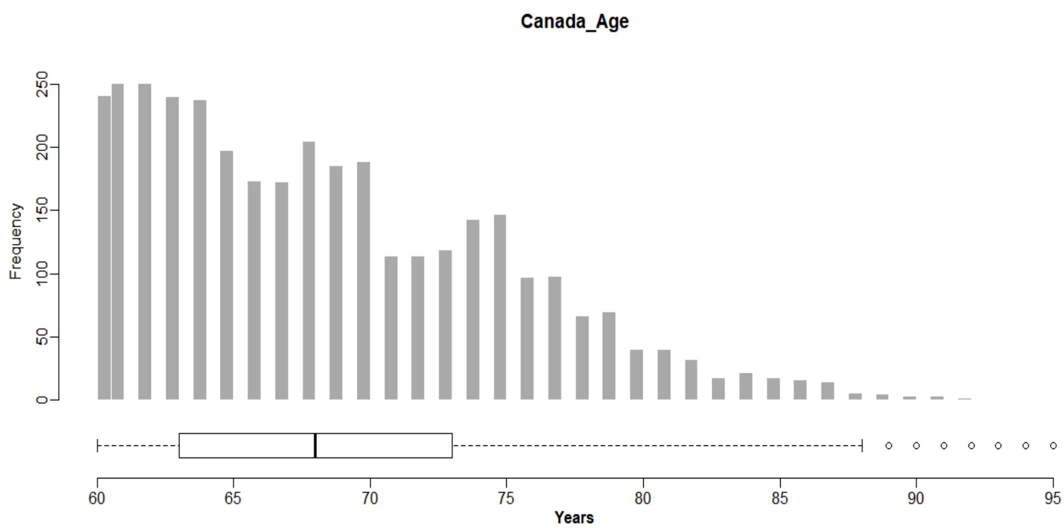


Figure II.2 Age distribution of Spain

Figure II.3 Age distribution of the Netherlands

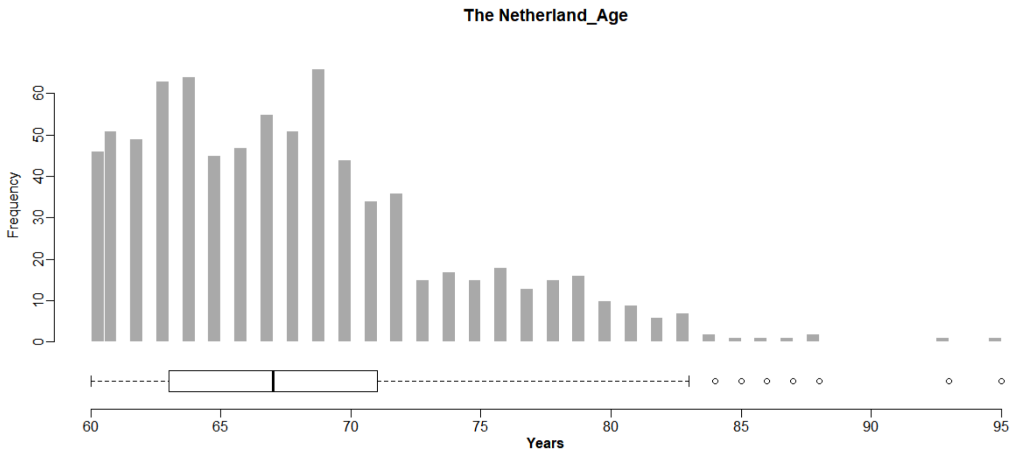


Table II. 3 Summary table of Age variable by country

Country	Min	1st Qu	Median	Mean	3rd Qu.	Max.
Canada	60.00	63.00	68.00	68.54	73.00	95.00
Spain	60.00	62.00	65.00	66.52	70.00	101.00
Netherlands	60.00	63.00	67.00	68.00	71.00	95.00

For reference, my tutor has provided me with the Z-score method of treatment of outliers, because it was the method applied in the same project from which database comes from.
"The Z-score, or standard score, is a way of describing a data point in terms of its relationship to the mean and standard deviation of a group of points"(Gorrie, 2016).

$$Z = \frac{X_i - \bar{X}}{sd}$$

Where X_ii ~ N (μ, σ2), and sd is the standard deviation of data.

The basic idea of this rule is that if X follows a normal distribution, N (μ, σ2), then Z. follows a standard normal distribution, N (0, 1), and Z-scores that exceed 3 in absolute value are generally considered as outliers. It presents a reasonable criterion for identification of the outlier when data follow the normal distribution. Since no z-score exceeds 3 in a sample size less than or equal to 10, the z-score method is not very good for outlier labeling. Another limitation of this rule is that the standard deviation can be inflated by a few or even a single observation having an extreme value. (Seo, 2002: 10).

The following table shows the outliers detected by the Z-score method.

Table II.4 Outliers detected by Z-score method

Country	Outliers in Age	Total Outliers
Canada	90 95 92 93 89 89 90 94 92 91 90 89 89 91 91 89	16
Spain	84 86 84 84 86 87 89 93 90 101 84 86 88 88	14
Netherlands	88 93 86 88 95 87	6

It has been found 16 outliers in Canada, 14 in Spain and 6 in the Netherlands. Table II. 5 gather the results after eliminating the outliers. Comparing with the results of TableII.3 it can be seen that the average age of each country has a slight decrease and the maximum value of each country has dropped from the original 95 (Canada), 101 (Spain), 95 (Netherlands) to 88, 83, 85 respectively. It should be noted that the values of minimum and quartiles remain the same after removing the outliers. **Appendix Figure A.1,** provides a comparison with boxplot before and after eliminating outliers.

Table II. 5 Summary table of after removing the outliers by Z-score method

Country	Min	1st Qu	Median	Mean	3rd Qu.	Max.
Canada	60.00	63.00	67.00	68.44	73.00	88.00
Spain	60.00	62.00	65.00	66.38	70.00	83.00
Netherlands	60.00	63.00	67.00	67.84	71.00	85.00

But I realized that there is a mistake in this method where the Age did not follow a normal distribution. **Appendix Figure A.2** shows the Shapiro normality test results where the p-values were lower than 0.05 so it cannot be accept the null hypothesis of normal distribution.

Therefore, I tried an alternative by applying the interquartile range method.

- Interquartile range (IQR) method

The interquartile range (IQR), also called the midspread or middle 50%, or technically H-spread, is a measure of statistical dispersion, being equal to the difference between 75th and 25th percentiles, or between upper and lower quartiles, IQR = Q3 − Q1. In other words, the IQR is the first quartile subtracted from the third quartile; these quartiles can be clearly seen on a box plot on the data. It is a trimmed estimator, defined as the 25% trimmed range, and is the most significant basic robust measure of scale. In R there is a function IQR() for computing the range value.

The interquartile range is often used to find outliers in data. Outliers here are defined as observations that fall below Q1 − 1.5 IQR or above Q3 + 1.5 IQR. In a boxplot, the highest and lowest occurring values within this limit are indicated by whiskers of the box (frequently with an additional bar at the end of the whisker) and any outliers as individual points. (Interquartile range (n.d))

Table II. 6 shows the detected outliers by applying the IQR method.

Table II. 6 Outliers detected by IQR method

Country	Outliers in Age	Total Outliers
Canada	90 88 95 92 93 88 88 89 88 89 90 94 92 91 88 88 90 89 89 91 91 89	22
Spain	83 84 82 83 86 84 83 84 86 82 83 87 82 82 89 83 83 82 83 93 90 82 101 83 82 84 86 88 82 82 83 88 82	33
Netherlands	83 88 93 84 86 88 83 83 83 83 85 95 83 83 87 84	16

Through the IQR method, Spain is the country with the highest number of outliers, which was 33, secondly 22 outliers were found in Canada and 16 in Netherlands. Comparing the TableII.7 with the initial data (TableII.3), the quartile values stay constant. The maximum values have fallen sharply, Spain dropped from the highest age of 101 to 81. At the same time, Canada's highest age dropped from 95 to 87 years after removing the outliers, while the Netherlands dropped from 95 to 82.

Table II. 7 Summary table of after removing the outliers by IQR method

Country	Min	1st Qu	Median	Mean	3rd Qu.	Max.
Canada	60.00	63.00	67.00	68.41	73.00	87.00
Spain	60.00	62.00	65.00	66.24	70.00	81.00
Netherlands	60.00	63.00	67.00	67.64	71.00	82.00

Comparing these two methods, the outliers that were detected with the IQR method include those of the Z-score method. However, the IQR method detects more outliers, such as finding 33 outliers in the Spanish age group, which is much more than the Z-score method. In contrast to the Netherlands, the IQR method removed the ages of 83, 84, and 85 years old but in the Z- score method these are not outliers.

This sample has to have fewer individuals in the older segment, as life expectancy is below 85 years old in the three countries. Therefore, among the older people is where may be more outliers despite they are legitimate part of the sample and the population under study (Tabachnick, Fidell, 2007). The IQR method penalizes this fact more than Z-score and it aims to eliminate individuals of more advanced ages, but we are interested in the population with older age and this extra deletion of cases may causes the loss of the wealth of data.

Thus after considerations I decided to use the results of the Z-score method for the subsequent analysis. Because it has the following advantages:

a) Keep the maximum number of possible observations of people in the group of 80 years and over.

b) I only perform an independence test in which age is involved and I do not run further estimations.

c) The results are not affected by the extreme values to obtain consistent conclusions because the age variable is used in the “Age_5cat” version (see Table II.1).

In this way, after removing the outliers the database turned to:

Table II.8 Sample size before and after remove the outliers

	Before	After
Canada	3,538	3,522
Spain	2,238	2,224
The Netherland	801	795

II.3.4 Weight cases of Canada

In the three countries, age and gender quotas were established to reach a representative sample of the older online population. In Canada, however, the final sample was unbalanced, in province of Ontario there are much more respondents than other provinces which was 1,374, almost 40% of the total sample of Canada. For this reason it was decided to weight the data of Canada according to province, gender and age and to maximize the representative nature of the final sample. Weights were devised using census data from Statistics Canada. The weight variable is included in the raw data file provided in SPSS format. The Table II.9 shows the unweighted and weighted results (Léger 2017).

After weighing the data, the sample size of Canada becomes 3,515 individuals.

Table II.9 Comparison of unweighted and weighted results of Canada (Image extracted from the Léger (2017:3))

	Total		
	Unweighted Sample	Weighted Sample	Weighted Sample (%)
Province			
British Columbia	408	**509**	**14%**
Alberta	406	**315**	**9%**
Manitoba/Saskatchewan	251	**223**	**6%**
Ontario	1,374	**1,343**	**38%**
Quebec	866	**892**	**25%**
Atlantic	255	**278**	**8%**
Gender			
Male	1,768	**1,649**	**46%**
Female	1,792	**1,911**	**54%**
Age			
Between 60 and 64	1,232	**990**	**28%**
Between 65 and 69	940	**853**	**24%**
Between 70 and 74	683	**616**	**17%**
75 or older	705	**1,101**	**31%**

II.4 Latent Class Analysis

II.4.1 Definition of the latent class analysis

Following McCutcheon (1987:04), Latent Class Analysis is a developing methodology for analysing categorical data. It enables a characterization of categorical latent (unobserved) variables from an analysis of the structure of the relationships among several categorical manifest (observed) variables. The method, which is often referred to as a “categorical data analogue to factor analysis”, was originally conceived of as an analytic method for survey data. As an exploratory technique, latent class analysis can be used to reduce a set of several categorically scored variables into a single latent variable with a set of underlying types or “classes”.

II.4.2 Characteristics of the latent class analysis

As Reyna and Brussino (2011: 13) point out, the LCA is based on the concept of probability and uses the observed data to estimate the parameters of the model: the probability of each latent class, whose sum must be equal to 1 (size); and the probabilities of conditional response, which represents the probability of a particular response in an observed variable conditioned by belonging to a certain latent class.
The cluster model of latent classes for mixed observed variables can be expressed as:

$$f(y_i|\theta) = \sum_{k=1}^{k} \pi_k \prod_{j=1}^{J} f_k\,(y_{ij}|\theta_{jk})$$

Where y_i represents the responses of a subject or object in a set of observed variables, k is the number of classes, π_k indicates the probability of belonging to a latent class k (size of class k). J indicates the total number of indicators and j a particular indicator, and $f_k(y_{ij}|\theta_{jk})$ implies the univariate distribution function of each of the elements y_{ij} of y_i, conditioned by the set of indicator variables j of class k. That is, the density function of a set of responses of a subject in a set of observed variables is equal to the sum of the probability of belonging to each of the classes by the product of the density function of each conditioned indicator for the class.

II.4.3 Model selection

One of the advantages of LCA is the variety of tools available to evaluate the fit of the model and determine the appropriate number of latent classes. Models with more parameters provide a better fit to the data, while models with fewer classes tend to have a worse fit, so the goal is to find the most parsimonious model that has an acceptable fit to the observed data (McCutcheon, 2002).

Following Beath (2017), the usual method used is an information criterion with the two main ones that are used being the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Using simulation that BIC is superior to AIC for selection in latent class models, much of the purely mathematical or Bayesian literature recommends BIC as a good indicator. With BIC the penalty is greater than for AIC and dependent on the number of observations, so will select models with a smaller number of classes. I used BIC for model selection, but other information criteria are provided too. Such as entropy, is a summary measure of classification quality based on the posterior probabilities that ranges from 0 to 1. (Ramaswamy, DeSarbo, Reibstein, & Robinson, 1993). In addition, CAIC, is quiet similar to BIC a consistent version of AIC, but penalizes more for model complexity than BIC.

There are a number of packages capable of fitting latent class models in R. poLCA and BayesLCA are two of these for fitting of latent class models. BayesLCA is particularly designed to perform Bayesian analyses, but has limited facilities for producing plots and summaries. poLCA is a more fully featured package which allows for polytomous outcomes and latent class regression (Beath, 2017:2). Therefore, I chose the poLCA package for the latent class analysis of this study.

II.5 R libraries

RStudio is the software used for all the analysis in this study. Then, the main functions used in each package will be explained in detail:

foreign: The function read.spss( ) was used to read SPSS data file in R.

survey: The function svydesign( ) specifying sampling weights.

poLCA: The function poLCA( ) estimates latent class and latent class regressions models for polytomous outcome variables. I chose maxiter=40000 that is the number of iterations through which the estimation algorithm will cycle. And nrep=10 it runs the model 10 times and keeps the model with the lowest BIC.

entropy: The function entropy( ) estimates the Shannon entropy H of the random variable Y from the corresponding observed counts y.

ggplot2: The function ggplot( ) was used to declare the input data frame for a graphic and to map variables to aesthetics.

reshape2: The function melt () convert an object into a molten data frame.

gmodels: The function CrossTable( ) implements a cross-tabulation with test for factor independence such as Chi-square test.

Car: The function Levenetest( ) for computing homogeneity of variance across groups, in this case for the variable “Q22_Age”.

III. OLDER AUDIENCE ANALYSIS: comparative of 3 countries

III.1 Univariate description analysis

III.1.1 Socio-demographic characteristics

The ratio of men respondents in Spain and the Netherlands is slightly higher than that of women (Table III. 1). However, it is more balanced in Canada; the sex ratio was approximately 50%. The age distribution is asymmetric in three countries; there are more individuals in the younger group and least in the group of 80 years and over (Table III.2).

Table III. 1 Gender distribution by country of the internet users aged 60 and over

	MALE	FEMALE	Total
CANADA	49.6%	50.4%	3,515
SPAIN	53.8%	46.2%	2,224
THE NETHERLANDS	52.7%	47.3%	795

Table III.2 Age distribution by country of the internet users aged 60 and over

	[60,65[	[65,70[	[70,75[	[75,80[	[80,+]
CANADA (N=3,515)	27.9%	24.2%	17.4%	21.4%	9.1%
SPAIN (N=2,224)	48,3%	22,1%	20.1%	7.0%	2.5%
THE NETHERLANDS (N=795)	34.3%	33.2%	18.4%	9.7%	4.4%

Regarding the education (Table III.3), secondary education is the most common educational level in Canada (56.2%) and the Netherlands (56.7%) while in Spain it was tertiary (39.8%), closely followed by secondary (37.4%). In Canada, about 97% of respondents have at least 10 years education while in Spain and the Netherlands are roughly 80%. This result shows that most respondents have received at least secondary education.

Table III.3 Education distribution by country

Education (%)	CANADA (N=3,515)	SPAIN (N=2,224)	THE NETHERLANDS (N=795)
Primary or less (up to 8-9 years)	2.4	21.8	12.2
Secondary (between 10 and 14 years)	56.2	37.4	56.7
Tertiary (15 years or more)	41.2	39.8	29.8
Don't know	0.2	1.0	1.3

The most common marital status in the sample of three countries is married, the ratio in Spain and the Netherlands was 75% and 72% respectively. Canada's married rate is relatively lower (65%) but it is still roughly two times more frequent than the category "Not married" (Table III.4). In addition, Table III.5 shows most of the respondents in the Netherlands and Spain claimed that they have descendants (75% in the Netherlands and 56% in Spain). However, the situation in Canada is the opposite where half of the respondents without children (53%) which is two times more frequent than the Netherlands (25%).

Table III.4 Marital status

Do you live with a partner?(%)	**CANADA** (N=3,515)	**SPAIN** (N=2,224)	**THE NETHERLANDS** (N=795)
Married	64.5	74.9	72.2
Not married	35.5	23.1	27.4
Prefer not to respond	0	1.98	0.4

Table III.5 Family status (Children/No children)

Do you have children? (%)	**CANADA** (N=3,515)	**SPAIN** (N=2,224)	**THE NETHERLANDS** (N=795)
Yes	46.5	56.3	75.0
No	53.5	41.7	24.6
Prefer not to respond	0	2.0	0.4

Personal income can be understood as an indicator of socioeconomic status (Table III.6), almost half of the samples in Spain (48%) and Canada (44%) claimed that their income were higher than the national average, and 31.6% of Canadian respondents are below the average while in Spain it was 11%. The situation in the Netherlands has changed. The most common income is lower than the national average income representing 32% of total sample.

Table III.6 Income

Income (%)	**CANADA** (N=3,522)	**SPAIN** (N=2,224)	**THE NETHERLANDS** (N=795)
Above the average	44.1	47.8	28.1
Similar to the average	11.2	17.1	16.0
Below the average	31.6	11.1	32.7
Not declared(Don't know or Prefer not to answer)	13.1	24.0	23.3

Regarding the employment rate, the normal retirement age is 65 in these countries (Table III.7). As expected about 70% of respondents in Spain and the Netherlands said that they are now retirees or in an unpaid position (e.g. housework). While in Canada, the ratio is higher and 80% of the sample are inactive. However, Canadian labour laws do not specify a retirement age and cannot force to retire according to age. When you reach 60, you become eligible to receive a reduced pension. In addition, nearly 30% of the respondents in Spain are still in employment, representing the highest ratio among the three countries.

Table III.7 Employment

Employment (%)	CANADA (N=3,522)	SPAIN (N=2,224)	THE NETHERLANDS (N=795)
Active (Full time work, part time work)	16.2	28.6	21.1
Inactive (Unemployed, retired or in unpaid positions)	80.3	70.2	72.0
Other	3.3	1.0	6.0
Not declared	0.1	0.3	0.9

III.1.2 Media & Internet usage

The uses of media in these three countries (Canada, Spain, The Netherlands) is quiet similar (Figure III.1), it has almost the same behaviour. The most prevalent format of mass media consumption is watching televisions on a TV set, about 90% percent of respondents in each country said they watched it the day before. The percentage of Canada is slightly lower than that of other countries, which is 87.62% with a difference of 5 percentual points from the Netherlands approximately. However, the ratings for watching TV on other devices are much lower, for example: watched TV on a computer (Spain 14.21%, the Netherlands 13.58%, Canada 8.81%), watched TV on a mobile (Spain 3.06% and Netherlands 2.26%, Canada 0.93%). This makes television and other mass media in stark contrast.

The second most commonly used media format is listening to the radio on a radio set, with more than half of the respondents stated that they used it in the day before (the Netherlands 64.91%, Spain 59.80% and Canada 56.65%). Similar to watching television, listening to the radio on a computer (Spain 11.38% and the Netherlands 10.19%, Canada 8.29%) or mobile phone (Spain 10.03% and the Netherlands 4.28%, Canada 2.44%) is much lower than conventional radios. The fact that televisions and radios occupy the first place in use indicates that non-Internet broadcast mass media dominate the elderly population.

Reading newspapers or magazines is also a popular media. In the Netherlands 69.07% of the respondents read newspapers or magazines in a printed version the day before (even higher than the use of radio 64.91%), then 44.03% of respondents chose to read on the internet. For Spain, more than half of the sample chose reading newspapers or magazines in the printed version (55.76%) and on the internet (50.76%). However, the frequency of reading on the Internet (41.35%) in Canada is lower than that of traditional paper reading (53.79%).

The frequency of reading books in traditional printed formats (approximately 43% in three countries) far exceeds the electronic version (24.19% in Spain, 17.61% in the Netherlands, 15.01% in Canada) and the audio version (4.96% in Spain, 1.89% in the Netherlands, Canada 1.18%).

Finally, Spain has slightly more media users than the Netherlands and Canada, while Canada has the least. In the elderly population, traditional mass media consumption marked difference with the broadcast media on others devices, non-Internet broadcast mass media occupy the upper hand.

Figure III. 1 Elderly media users

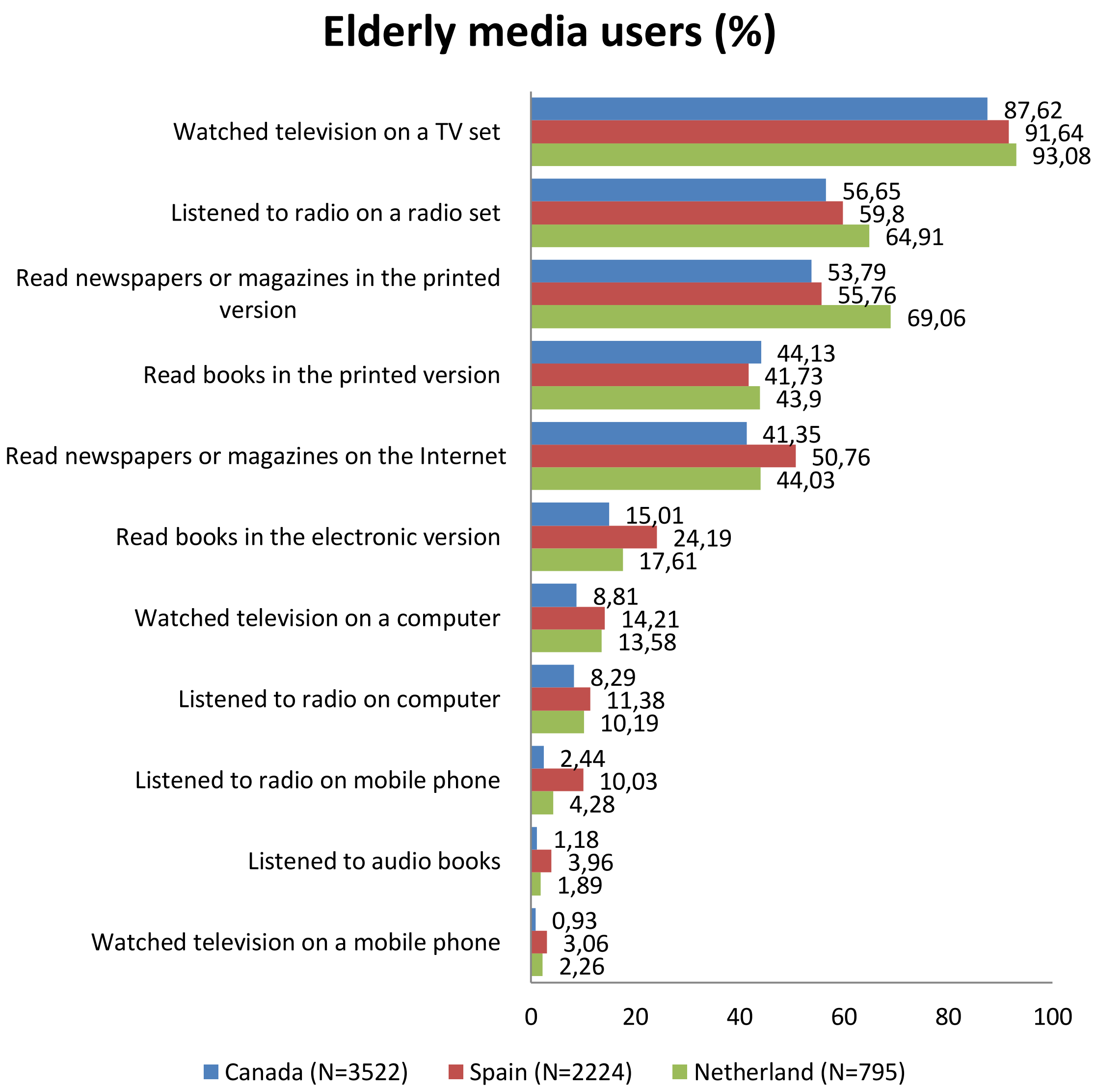
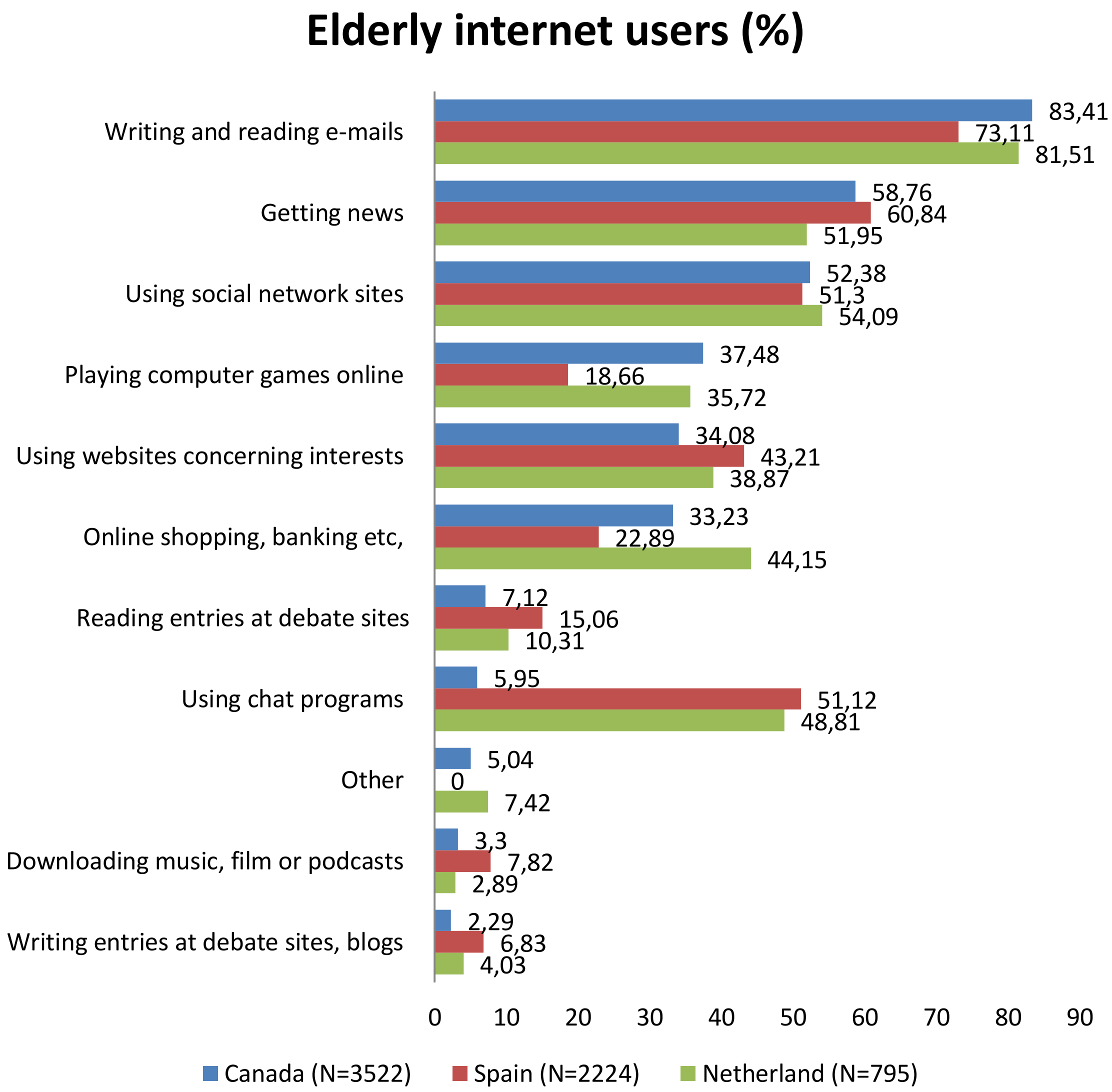


Figure III.2 shows, the most commonly used internet activities in the previous day were writing and reading e-mails occupied roughly 80% of the total sample of three countries. Then, getting news and using social network sites (e.g. Facebook) are the second popular internet activities. More than 50% of the sample used it the day before filling the survey.

For chat programs, about half of the sample in Spain and the Netherlands claimed that they used it the day before, but only 6% in Canada, which is a stark contrast. Similarly, one-third of respondents (about 35%) in Canada and the Netherlands stated that they had played computer games or shopped online the day before, but only 20% of respondents in Spain did it.

Another frequent activity is using websites concerning personal interests; about 40% of the sample in Spain selected this option, although in the case of Canada it was 34%. The least frequent activities are reading entries at debates sites (about 10%), downloading music, film or podcast (<10%) and writing entries at debates sites (<5%).

Figure III. 2 Internet users aged 60 and over by country



III.1.3 Mobile phone usage

The mobile phone seems to have become an indispensable communication tool. In the Dutch survey, almost 100% of elderly people have a mobile phone (Table III. 8) and in

Spain 90%. The mobile phone users in Canada is relatively lower, 18% of respondents did not have a mobile.

Table III. 8 Mobile phone usage

	Has mobile phone	Doesn't has mobile phone	Total Sample
Canada	81.7%	18.3%	3,515
Spain	90.4%	9.6%	2,224
The Netherland	96.4%	3.6%	795

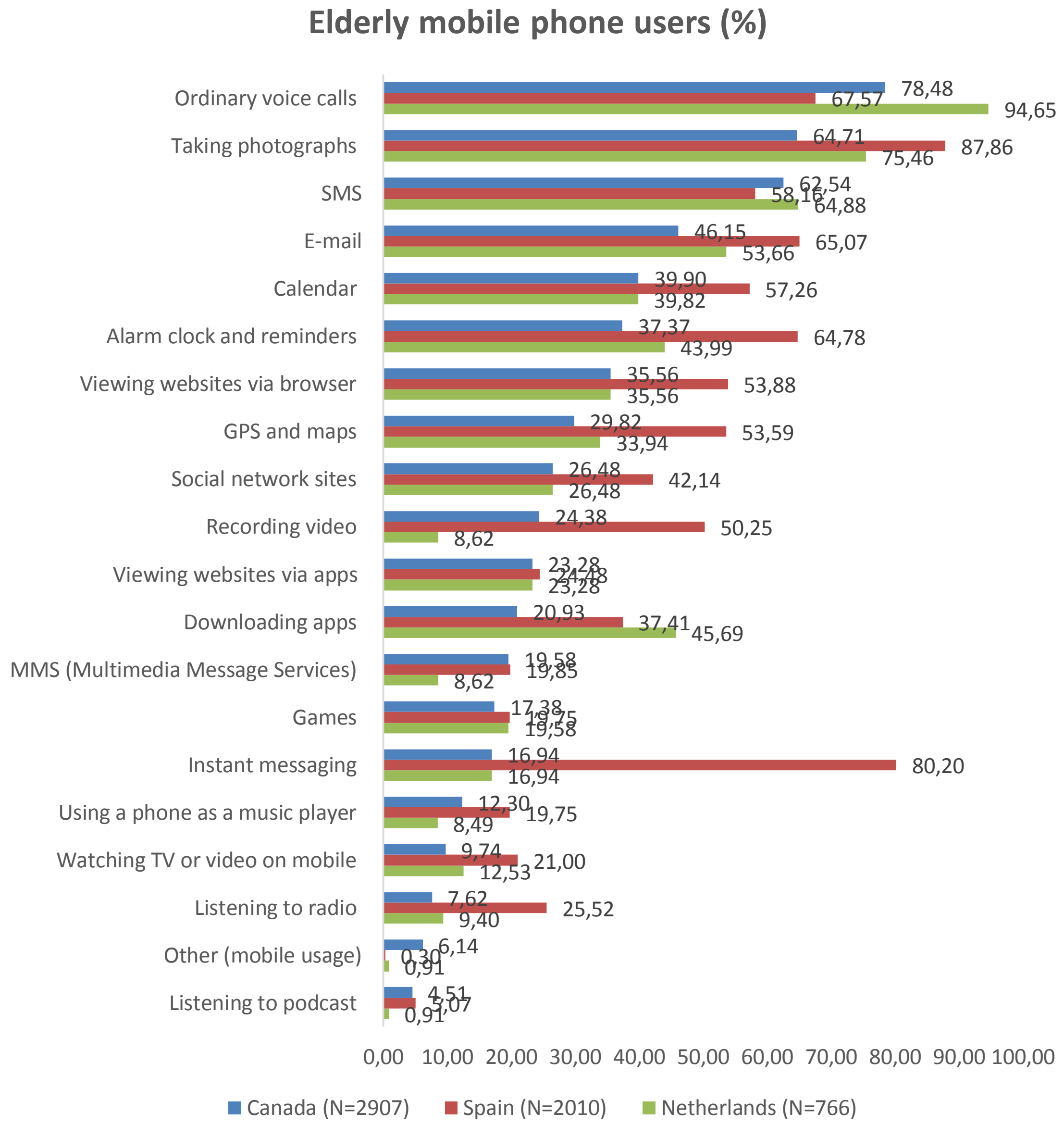
According the **Figure III. 3**, it should note that Spanish respondents use far more mobile phone features than other countries. However, Canadians and Dutch are very similar in their use of mobile features. In Spain, the most common mobile phone feature is taking photographs (88%) where in Canada and the Netherlands this is the second popular function. It is followed by instant messaging (e.g. Whatsapp) which was 80%. In stark contrast, respondents in Canada and the Netherlands rarely use this feature, only 17% of sample stated that they used it.

Users of email, alarm clock, reminders, Calendar, website browser, GPS & maps, recording video in Spain occupy more than half of the sample, almost 50% beyond other countries. Especially for video recording, 50% of respondents in Spain chose this option, while the Netherlands only has 24% and Canada has the lowest 9%.

In Canada and the Netherlands, the most practical mobile function is ordinary voice call (95% in Canada, 78% in the Netherlands) where in Spain this traditional function took the third place (68%). Then SMS is also a common feature that occupy roughly 60% in each country.

Finally, in regard to the least frequently used features are Games (around 20%), instant messaging for Canada and the Netherlands (17%), using the mobile as a music player (20% in Spain, 12% in Canada, 8% in the Netherlands), watching TV o video (21% in Spain,13% in the Netherlands and 10 in Canada), listening to radio (25% in Spain, 9% in the Netherlands and 8% in Canada), listening to podcast(5% in Spain, 4.5% in Canada and 1% in the Netherlands) or other.

Figure III. 3 Mobile phone features usage



III.2 Results latent class Analysis

I run two LCAs for each country, one for media with internet uses and another for mobile phone. I discarded using the 2 set of variables together in one LCA because there were both technical and analytical issues. Firstly, result was not satisfactory, it produced excessive classes and the information was mixed together. Secondly, media and internet usage gathered data on use the previous day. However, mobile phone usage refers to regular use of its features. As the meaning of the original variables is different, treating the 2 sets of variables of if they were equal seems to create problems.

In order to select variables that should be included in the latent class analysis, first I made the clusters with all the variables. Then I removed those variables that have less than 10% in the category "Yes" in the latent classes (all cluster variables are dichotomous), and in the descriptive analysis their population also was less than 10%.

The poLCA package of R does not allow analysis with weighted data, is still undergoing active development. Therefore, for Canada I used unweighted data to process the latent class analysis.

III.2.1 LCA for Canada

III.2.1.1 Media & Internet Uses

The exploratory analysis consisted on performing an LCA on 12 variables related to the media and internet usage in Canada. Table III. 9 shows the selected variables for the latent class analysis.

Table III. 9 Selected variables for media and internet cluster

Variables for media & internet cluster	**User(%)**
Watched television on a TV set	87,62
Writing and reading e-mails	83,41
Getting news	58,76
Listened to radio on a radio set	56,65
Read newspapers or magazines in the printed version	53,79
Using social network sites	52,38
Read books in the printed version	44,13
Read newspapers or magazines on the Internet	41,35
Playing computer games online	37,48
Using websites concerning my interests or hobbies	34,08
Online shopping, banking, travel reservation etc.	33,23
Read books in the electronic version	15,01
Watched television on a computer *	8,81
Listened to radio on computer *	8,29
Reading entries at debate sites, blogs *	7,12
Using chat programs *	5,95
Other *	5,04
Downloading music, film or podcasts *	3,3
Listened to radio on mobile phone *	2,44
Writing entries at debate sites, blogs *	2,29
Listened to audio books *	1,18
Watched television on a mobile phone *	0,93
N = 3,515. ***Variables removed from LCA**	

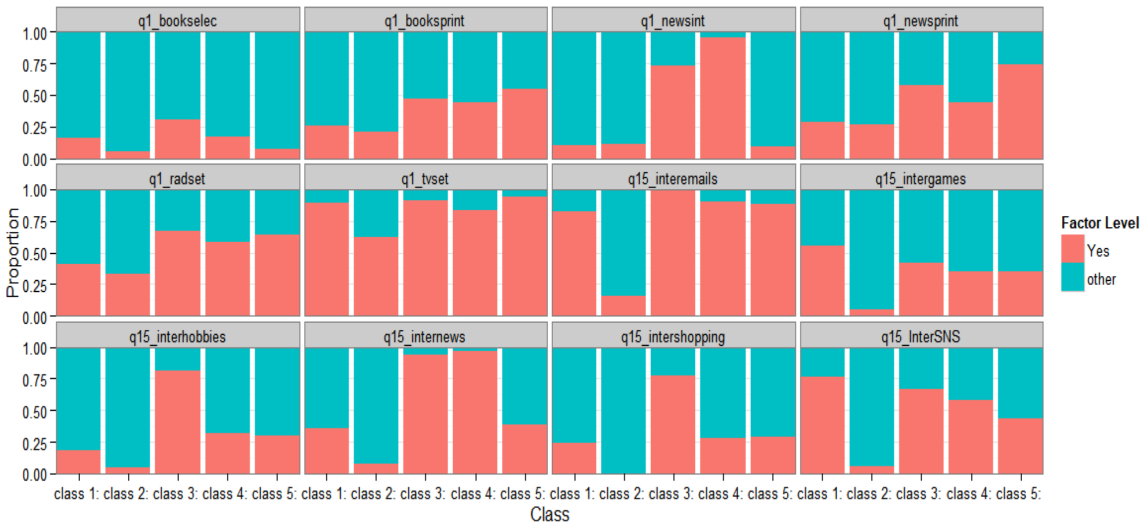
In order to determine the number of classes, I ran several solutions ranging from 3 to 7 classes (Table III.10). According to the BIC the model 5 (five classes) has been chosen as having the lowest BIC (44995.70) and CAIC (45099.70), also a high entropy (0.79).

Table III. 10 Fit statistics

	Modell	log_likelihood	df	BIC	CAIC	Entropy
1	Model 3	-26434.45	3478	53228.24	53272.24	0.674
2	Model 4	-26338.65	3463	53159.14	53218.14	0.753
3	Model 5	-22083.16	2803	44995.70	45099.70	0.793
4	Model 6	-26229.31	3433	53185.46	53274.46	0.850
5	Model 7	-26171.16	3418	53191.67	53295.67	0.599

Figure III.4 shows the behavior of classes of model 5 and **Appendix Table.A.3** contains the detailed description tables and graphs of each class.

Figure III. 4 GG plot of the model 5



Cross table was used to study the socio-demographic characteristics of each class. Moreover, according the chi-square test there are no association between the classes with the variable "has_children" (see the evidence of relationship between the indicated variables in **Appendix Table.A.4**). Therefore, this variable is not significant for the description of the clusters. **Appendix Table A.5** gathers all the outputs of the cross tables which present the socio-demographic characteristics of the 5 classes. Results can be summarized as follow: I only highlight the most relevant characteristics of each clusters.

- Class CI-1 (14.9%): Dominates watching TV on a TV set (87.4%), using social network sites (88.2%), playing computer games online (60.1%). Mainly dominated by women, unmarried people (single, divorced or widowed), who has income below the average and inactive in the labor sector.

- Class CI-2 (9.8%): Compared to other classes, this class contains people who use the media and internet less frequently, because all the variables take the lowest values in this class. Predominate old people who are 75 to 80+ years old, lower education levels (primary or less), prefer not to declared the income and with inactive status.

- Class CI-3 (15.6%): Dominates watching TV on a TV set (90.4%), listening to the radio on a radio set (67.9%), reading books in electronic version (36.07%), writing and reading email (99.8%), getting news on internet(94.2%), using social network sites (66.9%), online shopping, banking (93,08%) and using web sites concerning interests (95.8%). Predominate seniors aged 60 to 70, in a married state, has a higher level of education (tertiary), income above the average and employed.

- Class CI-4 (26.8%): Dominates getting news on internet (100%), writing and reading email (89.8%). Mainly dominated by men, old people who are 60 to 70 years old, in a married state and a higher level of education (tertiary).

- Class CI-5 (33.6%): Dominates watching TV on a TV set (94.7%), reading newspapers in printed version (78.2%), reading books on printed version (56.8%), listening to the radio on a radio set (66.1%) and writing and reading email (90.2%). Predominate people who are aged 75 to 80+.

III.2.1.2 Mobile phone uses

Among the total set of 16 variables that report the use of mobile phone, Table III. 11 shows the selected variables for the latent class analysis of mobile phone usage.

Table III. 11 Selected variables for mobile phone cluster

	Users(%)		**Users(%)**
Ordinary voice calls	78.48	Viewing websites via apps	23.28
Taking photographs	64.71	Downloading apps	20.93
SMS	62.54	MMS (Multimedia Message Services)	19.58
E-mail	46.15	Games	17.38
Calendar	39.9	Instant messaging (WhatsApp, etc.)	16.94
Alarm clock and reminders	37.37	Using a phone as a music player	12.3
Viewing websites via browser	35.56	Watching TV or video on mobile *	9.74
GPS and maps	29.82	Listening to radio *	7.62
Social network sites	26.48	Other (mobile usage) *	6.14
Recording video	24.38	Listening to podcast *	4.51
N=2,907. Question shown if ownership of mobile phone was selected. ***Variables removed from LCA**			

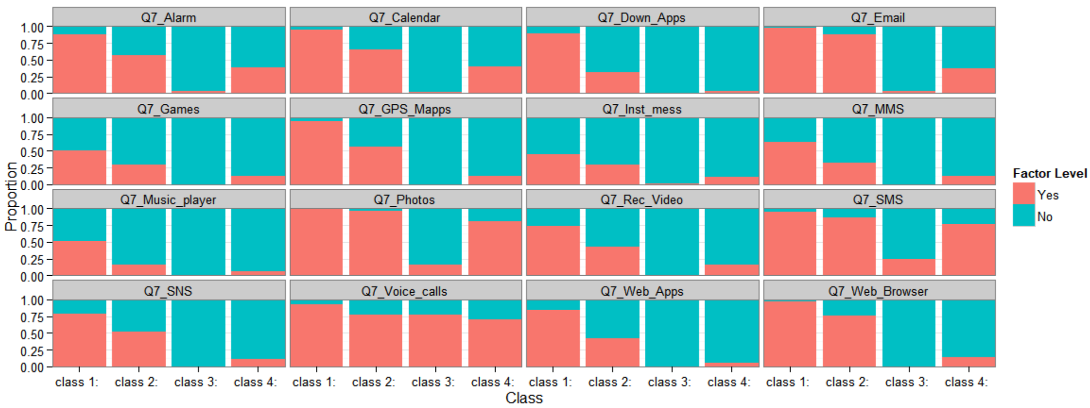
After an exploration of the results, I decided to select the solution of 4 classes (Model 4) although the model 4 has slightly higher BIC than model 5 but it has a higher entropy and slightly lower CAIC which penalizes more for model complexity than BIC. In addition, according to the descriptive analysis, the model 4 represents the same behavior with model 5 but has fewer clusters (Table III.12).

Table III. 12 Fit statistics

	Modell	log_likelihood	df	BIC	CAIC	Entropy
1	Model 3	-20314.10	2857	41026.94	41076.94	0.860
2	Model 4	-20101.13	2840	40736.58	40803.58	0.799
3	Model 5	-20027.46	2823	40724.80	40808.80	0.762
4	Model 6	-19968.16	2806	40741.78	40842.78	0.724
5	Model 7	-19920.21	2789	40781.45	40899.45	0.720

Figure III.4 shows the behavior of classes of model 4 and **Appendix Table.A.6** contains the detailed description tables and graphs of each class.

Figure III. 4 GG plot of the model 4



Cross table was used to study the socio-demographic characteristics of each class. Moreover, according the chi-square test there are no association between the classes with the variable the variable "has_partner" and "has_children" as having the p-value greater than 0.05 that accept the null hypothesis (see the evidence of relationship between the indicated variables in **Appendix Table.A.4**). Therefore, those variables are not significant for the description of the clusters. **Appendix Table A.7** gathers all the outputs of the cross tables which present the socio-demographic characteristics of the 4 classes. Results can be summarized as follow:

Class 1 (15.2%): Contrasting other classes, the use of each mobile feature in this class is the highest. Mainly dominated by men, seniors aged 60 to 65, with high education level (tertiary), has income above the average and active in the labor sector.

Class 2 (26.0%): In class 2 it is also used in every mobile phone function but lower than class 1 and is the second most used class. Predominate old people who are 65 to 70 years old, with high education level (tertiary), has income above the average or prefer not to declared and with active status.

Class 3 (29.9%): Predominates the function voice call, then the use of other functions is almost zero. Mainly dominated by women, aged 75 to 80, has medium level of education (secondary), income below the average and unemployed (retired or in unpaid position like housework).

Class 4 (28.9%): People in this class don't often use mobile phone, the percentages are below the weighted average. Predominate old people who are 70 to 75 years old, with lower level of education (primary or less) and has income similar to average or prefer not to declare.

III.2.1.3 Cross table of Internet and media usage & mobile phone usage

After this, I decided to cross the classes of media and internet usage with the classes of mobile phone usage to find out the relationship between them. The chi-square test shows the two variables are not independent while the p value was quasi zero (see the evidence of relationship between the indicated variables in **Appendix Table.A.4).**

According the Table III.13, the CM-1 is more related to CI-3 that means those who use more internet and media also use more mobile, vice versa. CM-2 is related to CI-4, old people who use less frequently the mobile phones use internet to getting news, writing and reading emails. CM-3 is related to CI-2, people who only use cell phones to call, hardly use the Internet and media. CM-4 is related to CI-5, old people who use traditional broadcast media (watching TV on a TV set, listening to the radio on a radio set, read news in printed version, etc.) often use less Internet and mobile phones.

Table III. 13 Cross table of CI & CM

(N=2871)	**CI-1**	**CI-2**	**CI-3**	**CI-4**	**CI-5**	**Column total**
CM-1	13.3%	5.0%	27.2%	15.8%	7.9%	13.6%
CM-2	22.6%	22.0%	27.1%	28.7%	21.7%	24.6%
CM-3	36.7%	44.1%	19.6%	29.9%	37.9%	33.2%
CM-4	27.4%	28.9%	26.2%	25.6%	32.5%	28.6%
TOTAL	100%	100%	100%	100%	100%	100%
*CI= Cluster of media & internet usage, CM= Cluster of mobile phone usage						

III.2.2 LCA for Spain

III.2.2.1 Media & Internet Uses

Among the total set of 16 variables that report the use of media and internet, Table III. 14 shows the selected variables for the latent class analysis.

Table III. 14 Selected variables for media & internet cluster

Variables for media & internet cluster	**User(%)**

Watched television on a TV set	91,64
Writing and reading e-mails	73,11
Getting news	60,84
Listened to radio on a radio set	59,8
Read newspapers or magazines on the Internet	55,76
Using social network sites	51,3
Using chat programs	51,12
Read newspapers or magazines in the printed version	50,76
Using websites concerning my interests or hobbies	43,21
Read books in the printed version	41,73
Read books in the electronic version	24,19
Online shopping, banking, travel reservation etc,	22,89
Playing computer games online	18,66
Reading entries at debate sites, blogs	15,06
Watched television on a computer	14,21
Listened to radio on computer	11,38
Listened to radio on mobile phone *	10,03
Downloading music, film or podcasts *	7,82
Writing entries at debate sites, blogs *	6,83
Listened to audio books *	3,96
Watched television on a mobile phone *	3,06
Other *	0,18
N = 2,224. ***Variables removed from LCA**	

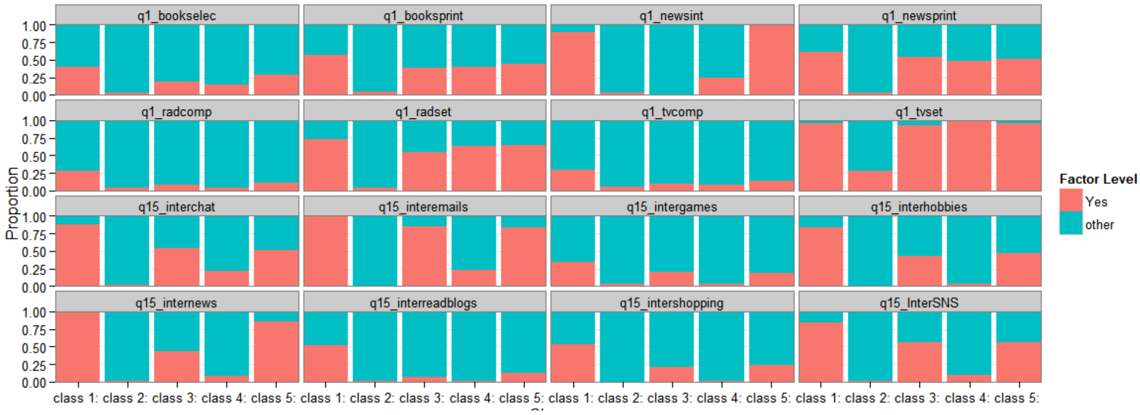
A cluster of five classes has been chosen as having the lowest BIC, CAIC and the highest entropy (Table III.15).

Table III. 15 Fit statistics

```
  Modell  log_likelihood  df      BIC      CAIC       Entropy
1 Model 3     -18719.610 2174 37824.57 37874.57        0.711
2 Model 4     -18618.160 2157 37752.69 37819.69        0.681
3 Model 5     -18509.047 2140 37665.49 37749.49        0.779
4 Model 6     -18457.481 2123 37693.38 37794.38        0.698
5 Model 7     -18416.961 2106 37743.35 37861.35        0.673
```

Figure III.5 shows the behavior of classes and **Appendix Table.A.8** contains the detailed description tables and graphs of each class of model 5.

Figure III. 5 GG plot of the model 5



Cross table was used to study the socio-demographic characteristics of each class. Moreover, according the chi-square test the classes are dependent with all the socio-demographic variables (see the evidence of relationship between the indicated variables in **Appendix Table.A.9**). **Appendix Table.A.10** gathers all the outputs of the cross tables which present the socio-demographic characteristics of the 5 classes. Results can be summarized as follow: I only highlight the most relevant characteristics of each clusters.

Class CI-1 (14.5%): Dominates all the kind of media and internet usage. Mainly dominated by seniors aged 60 to 65, have higher education (tertiary), income above the average and employed.

Class CI-2 (5.8%): Compared to other classes, this class contains people who rarely use the media and internet, because all the variables take the lowest values in this class. Mainly dominated by women, old people who are 65 to 70 years old, lower education levels (primary or less), have children and prefer not to declare the income.

Class CI-3 (26.8%): Dominates watching TV on a TV set (92.8%), read news in a printed version (67.9%), writing and reading email (85.4%), using social network sites (58.3%) and playing games online (21.4%). Predominated by old people with medium level of education (secondary) and unemployed (retired, in unpaid position).

Class CI-4 (26.8%): Dominates watching TV on a TV set (98.3%). Mainly dominated by women, old people who are 75 years old and above, unmarried state (single, divorced or widowed) without children, lower level of education (primary or less) and income below the average and unemployed.

Class CI-5 (33.6%): Dominates watching TV on a TV set (96.2%), reading books on printed version (43.2%), listening to the radio on a radio set (64.8%) and reading newspapers in electronic version (100%). Predominated by men, people who are aged 70 to 75 in married state with higher level of education (tertiary) and has income above the average.

III.2.2.2 Mobile phone Uses

Among the total set of 18 variables that report the use of mobile phone, Table III. 16 shows the selected variables for the latent class analysis of mobile phone usage.

Table III. 16 Selected variables for mobile phone cluster

	Users(%)		**Users(%)**
Taking photographs	87.86	Social network sites	42.14
Instant messaging (WhatsApp, etc.)	80.2	Downloading apps	37.41
Ordinary voice calls	67.56	Listening to radio	25.52
E-mail	65.07	Viewing websites via apps	24.48
Alarm clock and reminders	64.78	Watching TV or video on mobile	21.00
SMS	58.16	MMS (Multimedia Message Services)	19.85
Calendar	57.26	Using a phone as a music player	19.75
Viewing websites via browser	53.88	Games	19.75
GPS and maps	53.59	Listening to podcast *	5.07
Recording video	50.25	Other (mobile usage)*	0.3
N=2,010. Question shown if ownership of mobile phone was selected. ***Variables removed from LCA**			

After comparing the model 5 with model 6, a cluster of five classes has been chosen. Although it has slightly higher BIC (with a difference of 3,15) than 6 clusters, its has a lower CAIC which penalizes more for the model complexity and also a higher entropy than model 6. Then observing the descriptive results, the two models follow the same pattern. According to the parsimony principle, the model with lower number of clusters is better.

Table III. 17 Fit statistics

	Modell	log_likelihood	df	BIC	CAIC	Entropy
1	Model 3	-18180.25	1954	36786.43	36842.43	0.805
2	Model 4	-18018.95	1935	36608.34	36683.34	0.772
3	Model 5	-17927.63	1916	36570.22	36664.22	0.744
4	Model 6	-17853.80	1897	36567.07	36680.07	0.717
5	Model 7	-17801.26	1878	36606.51	36738.51	0.728

Figure III.6 shows the behavior of classes and **Appendix Table.A.11** contains the detailed description tables and graphs of each class of model 5.

Figure III. 6 GG plot of the model 5

Cross table was used to study the socio-demographic characteristics of each class. Moreover, according the chi-square test there are no association between the classes with the variable “Q21_Sex”, “has_partner” and “has_children” as having the p-value greater than 0.05 that accept the null hypothesis (see the evidence of relationship between the indicated variables in **Appendix Table.A.9**). Therefore, those variables are not significant for the description of the clusters. **Appendix Table.A.12** gathers all the outputs of the cross tables which present the socio-demographic characteristics of the 3 classes. Results can be summarized as follow:

Class CM-1 (20.8%): Predominates SMS and MMS (sending images or sound) is the second most used class of mobile functions but lower than class 4. Mainly dominated seniors aged 60 to 65, with high education level (tertiary).

Class CM-2 (21.2%): Predominates the function take photos, instant messaging (whatsapp), other functions are rarely used. Predominate old people who are 65 to 70 years old and has income above the average.

Class CM-3 (28.6%): People in this class do not often use mobile phone, the percentages are below the weighted average. Predominate old people with medium level of education (secondary) and unemployed.

Class CM-4 (21.5%): Contrasting other classes, the use of each mobile feature in this class is the highest, except SMS and MMS. Predominate old people who are 70 years old and over, with higher level of education (tertiary), has income above the average and active in the labor sector.

Class CM-5 (8.0%): Predominates voice call and others functions are rarely used.
Mainly dominated seniors aged 70 to 75 with lower level of education (primary or less) and income similar or below the average, while this is the cluster with the higher proportion of participant who did not declare their income.

III.2.2.3 Cross table of Internet and media usage & mobile phone usage

After this, I crossed the classes of media and internet usage with the classes of mobile phone usage to find out the relationship between them. The chi-square test shows the two variables are not independent while the p value was quasi zero (see the evidence of relationship between the indicated variables in **Appendix Table.A.9**).

According the Table III.18, the CM-1 is more related to CI-5 that means those who use conventional mass broadcast media (rarely use the internet) used to send SMS and MMS (images and sounds) on the mobile phone. CM-2 is related to CI-3, old people who use the traditional mass broadcast and do some internet activities use mobile phone to table photos and send instant messages (mainly whatsapp). CM-3 is related to CI-2, old people who rarely use media and the Internet rarely use mobile phones, vice versa. CM-4 is related to CI-1, older people who use the media and internet frequently have better use of mobile phone features. CM-5 is related to CI-4, old people who only watch TV on the TV set use the mobile phone as a tool for making calls.

Table III. 18 Cross table of CI & CM

(N=2010)	**CI-1**	**CI-2**	**CI-3**	**CI-4**	**CI-5**	**Column total**
CM-1	22.3%	15.5%	21.5%	3.7%	23.6%	20.7%
CM-2	17.8%	11.7%	24.5%	14.9%	23.6%	21.2%
CM-3	9.1%	49.5%	30.0%	44.4%	26.1%	38.6%
CM-4	50.5%	4.9%	15.0%	6.1%	21.8%	21.5%
CM-5	0.3%	18.4%	7.1%	23.8%	5.0%	8.0%
TOTAL	100%	100%	100%	100%	100%	100%
*CI= Cluster of media & internet usage, CM=Cluster of mobile phone usage						

III.2.3 LCA for the Netherlands

III.2.3.1 Media & Internet Uses

Among the total set of 16 variables that report the use of media and internet, Table III. 19 shows the selected variables for the latent class analysis.

Table III. 19 Selected variables for media & internet cluster

Variables for media & internet cluster	**User(%)**
Watched television on a TV set	93,08
Writing and reading e-mails	81,51
Read newspapers or magazines in the printed version	69,06
Listened to radio on a radio set	64,91
Using social network sites	54,09
Getting news	51,95
Using chat programs	48,81
Online shopping, banking, travel reservation etc,	44,15

Read newspapers or magazines on the Internet	44,03
Read books in the printed version	43,9
Using websites concerning my interests or hobbies	38,87
Playing computer games online	35,72
Read books in the electronic version	17,61
Watched television on a computer	13,58
Reading entries at debate sites, blogs	10,31
Listened to radio on computer	10,19
Other *	7,42
Listened to radio on mobile phone *	4,28
Writing entries at debate sites, blogs *	4,03
Downloading music, film or podcasts *	2,89
Watched television on a mobile phone *	2,26
Listened to audio books *	1,89

In this case, model 2 (2 classes) has been chosen as having the lowest BIC and CAIC (Table III. 20).

Table III. 20 Fit statistics

	Modell	log_likelihood	df	BIC	CAIC	Entropy
6	Model 2	-6427.249	764	13061.53	13092.53	0.550
1	Model 3	-6376.086	748	13066.05	13113.05	0.556
2	Model 4	-6342.187	732	13105.11	13168.11	0.569
3	Model 5	-6312.451	716	13152.49	13231.49	0.590
4	Model 6	-6292.488	700	13219.42	13314.42	0.633
5	Model 7	-6276.828	684	13294.95	13405.95	0.662

Figure III.7 shows the behavior of classes and **Appendix Table.A.13** contains the detailed description tables and graphs of each class of model 2.

Figure III. 7 GG plot of the model 2

Cross table was used to study the socio-demographic characteristics of each class. Moreover, according the chi-square test there are no association between the classes with the variable "Q22_Age_5cat", "has_children", "edu_3cat", "income_3cat" and "employ_3cat" (see the evidence of relationship between the indicated variables in **Appendix Table.A.14**). Therefore, those variables are not significant for the description of the clusters. **Appendix Table.A.15** gathers all the outputs of the cross tables which present the socio-demographic characteristics of the 2 classes. Results can be summarized as follow: I only highlight the most relevant characteristics of each clusters.

- Class CI-1 (46.0%): Contrasting with Class CI-2, the use of media and internet in this class is the highest. Moreover, predominate by men, in a married state and with higher level of education (tertiary).

- Class CI-2 (54.0%): The use of media and internet in this class is less frequently than the Class CI-1, in general the variables take the lower values than the weighted average. In addition, mainly dominated by women, in unmarried state (single, divorced or widowed) and with lower level of education (primary or less).

III.2.3.2 Mobile phone Uses

Among the total set of 14 variables that report the use of mobile phone, Table III. 21 shows the selected variables for the latent class analysis of mobile phone usage.

Table III. 21 Selected variables for mobile phone cluster

	Users (%)		**Users (%)**
Ordinary voice calls	94.65	Viewing websites via apps	23.28
Taking photographs	75.46	Games	19.58
SMS	64.88	Instant messaging (WhatsApp, etc.)	16.94
E-mail	53.66	Watching TV or video on mobile	12.53
Downloading apps	45.69	Listening to radio *	9.4
Alarm clock and reminders	43.99	Recording video *	8.62
Calendar	39.82	MMS (Multimedia Message Services) *	8.62
Viewing websites via browser	35.56	Using a phone as a music player *	8.49
GPS and maps	33.94	Other (mobile usage)*	0.91
Social network sites	26.48	Listening to podcast *	0.91
N=766. Question shown if ownership of mobile phone was selected. ***Variables removed from LCA**			

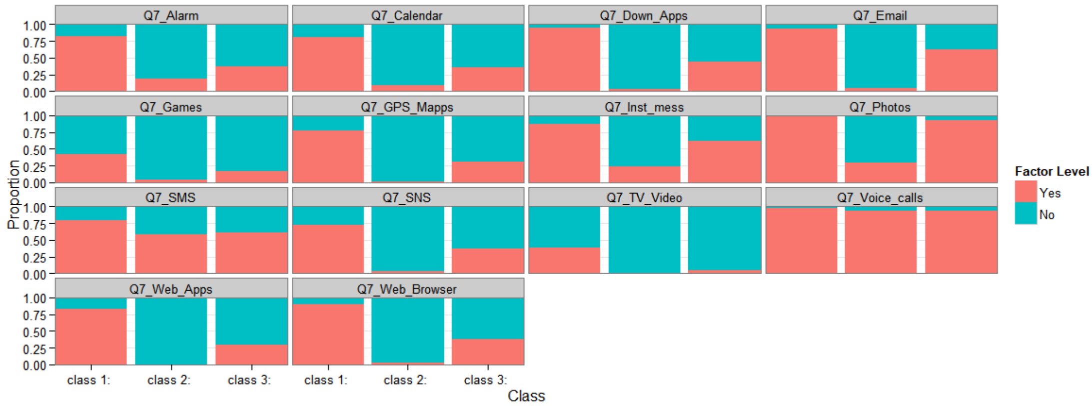
A cluster of three classes has been chosen as it has the lowest BIC, CAIC and high entropy (Table III.22).

Table III. 22 Fit statistics

	Modell	log_likelihood	df	BIC	CAIC	Entropy
6	Model 2	-5421.278	737	11035.15	11064.15	0.848
1	Model 3	-5235.806	722	10763.82	10807.82	0.816
2	Model 4	-5208.358	707	10808.55	10867.55	0.797
3	Model 5	-5182.081	692	10855.61	10929.61	0.730
4	Model 6	-5165.632	677	10922.33	11011.33	0.721
5	Model 7	-5151.065	662	10992.81	11096.81	0.749

Figure III.8 shows the behavior of classes and **Appendix Table.A.16** contains the detailed description tables and graphs of each class of model 3.

Figure III. 8 GG plot of the model 3



Cross table was used to study the socio-demographic characteristics of each class. Moreover, according the chi-square test there are no association between the classes with the variable “Q21_Sex”, “has_children” and “edu_3cat” as having the p-value greater than 0.05 that accept the null hypothesis (see the evidence of relationship between the indicated variables in **Appendix Table.A.14**). Therefore, those variables are not significant for the description of the clusters. **Appendix Table A.17** gathers all the outputs of the cross tables which present the socio-demographic characteristics of the 3 classes. Results can be summarized as follow:

- Class CM-1 (27.0%): Contrasting other classes, the use of each mobile feature in this class is the highest. Mainly dominated seniors aged 60 to 65, in a married state, have income above the average and employed.

- Class CM-2 (30.2%): Compared to other classes, this class contains people who rarely use the mobile phone, because all the variables take the lowest values in this class. Predominate old people aged 70 and over, not married, have income below the average or prefer not to declare and unemployed.

- Class CM-3 (42.8%): This class is the second most used class. The use of mobile phone features is slightly higher than the weighted average. Predominate old people aged 65 to 70 and have income similar to average.

III.2.3.3 Cross table of Internet and media usage & mobile phone usage

After this, I crossed the classes of media and internet usage with the classes of mobile phone usage to find out the relationship between them. The chi-square test shows the two variables are not independent while the p value was quasi zero (see the evidence of relationship between the indicated variables in in **Appendix Table.A.14**).

According to the Table III.23, CI-1 is more related to CM-1 and CI-2 with CM-2 which meanings older people who use the media and internet frequently have better use of mobile phone, also who rarely use media and the Internet rarely use mobile phones.

Table III. 23 Cross table of CI & CM

(N=766)	**CI-1**	**CI-2**	**Column total**
CM-1	38.5%	16.9%	27.0%
CM-2	19.0%	40.0%	30.2%
CM-3	42.5%	43.1%	42.8%
TOTAL	100%	100%	100%
*CI= Cluster of media & internet usage, CM=Cluster of mobile phone usage			

IV. CONCLUSION

This study discusses the results of an online study conducted in Canada, Spain and the Netherlands in 2016. Data come from the project "Cross-National Longitudinal Study: Older Audiences in the Digital Media Environment" that targeted older Internet users who are 60 years old and over, with no upper threshold on age.

Diversity of digital uses among older individuals. In the three countries, it is confirmed that socio-economic characteristics do shape the digital uses. As expected, individuals with higher income and higher educational level show higher use of digital tools in their everyday life. In addition, the internet use is positively associated with mobile phone use; older people who use the media and internet frequently show heavier use of the mobile phone.

Media and internet usage is similar in the 3 countries. Regarding to media, watching TV on a conventional TV set is the most popular activity in the 3 countries (87.6% in Canada, 91.6% in Spain and 93.1% in the Netherlands). Followed by listening to the radio on a radio set (the Netherlands 64.91%, Spain 59.80% and Canada 56.65%). Regarding to internet activities, reading or writing emails stand out as the most popular activity in the 3 countries (roughly 80% of the total sample in three countries). Followed by getting news and using social network sites (e.g. Facebook), more than 50% of the sample used it the day before filling the survey.

The country where the mobile phone is most popular is the Netherlands 96.4% of the respondents have mobile phone and in Spain 90%. The mobile phone users in Canada is relatively lower, 18% of respondents did not have a mobile. Elderly people use their cell phones to make voice calls (the Netherlands 94.7%, Canada 78.5% and in Spain 67.6%) and take photos (Spain 87.9%, the Netherlands 75.5% and Canada 64.7%). Moreover, in Spain the use of mobile phone features appears to be more intensive and more variegated than other countries. For example, the use of instant messaging (mainly Whatsapp) was 80% among the Spanish seniors. In stark contrast with Canadians and Dutch who rarely use this feature (roughly 17%).

In the elderly population, traditional mass media consumption marked difference with the broadcast media on others devices, predominates non-Internet broadcast mass media. The results also indicate that old people use the Internet for personal communication and information gathering, while the behaviour to satisfy entertainment needs is not significant. In general, old people who are in the younger age group (60-70) tend to have a more diverse use of internet and the mobile phone. There are differences in the Internet use of older people by gender, and marital status is also a factor that shapes the use of the Internet.

An in-depth understandingg of the Internet use of the elderly can provide the elderly with high-quality electronic services, thereby improving the quality of life. The contribution of this study is to discover through two latent class analysis the knowledge behind the data that reflects the Internet use of the elderly. Despite sample design aimed at granting representativeness of the older internet users in 3 countries, all the data was collected by means of a marketing panel. A computer assisted telephone interview (CATI) survey would help to test the stability of obtained results.

REFERENCES

Beath, K. J. (2017). randomLCA: An R Package for Latent Class with Random Effects Analysis. *Journal of Statistical Software*, *81*(13), 1-25

CTV News (2017): *https://www.ctvnews.ca/sci-tech/internet-usage-growing-fastest-among-older-canadians-statscan-1.3676986*

Gorrie, C. (2016) Three ways to detect outliers. Retried from *http://colingorrie.github.io/outlier-detection.html#z-score-method*

Interquartile range (n.d.). In Wikipedia. Retrieved June 15, 2018, from *https://en.wikipedia.org/wiki/Interquartile_range*

Léger (2017): Technical report – Canada. Older Media Audiences (First Wave). (5 pages)

Linzer DA, Lewis JB. poLCA: An R package for polytomous variable latent class analysis. Journal of Statistical Software.

Loos, E., Nimrod, G. and Fernández-Ardèvol, M. (2018) Older audiences in the digital media environment: A cross-national longitudinal study. Project Report. ACT Project, Montreal. Retrieved from *https://spectrum.library.concordia.ca/983866/*

McCutcheon AL. Basic concepts and procedures in single and multiple group latent class analysis. En: Hagenaars JA, McCutcheon AL, editores. Applied Latent Class Analysis. New York: Cambridge University Press; 2002. pp. 56-88.

McCutcheon, A.L.(1987) Latent class analysis (Sage University Paper series on Quantitative Applications in the Social Sciences, No. 07-064). Newbury Park, CA:Sage.

UOC News (2017): *https://www.uoc.edu/portal/es/news/actualitat/2017/221-dia-internacional-personas-mayores.html*

Pulletikurtt, H. (2015) Everything about data science. Retrieved from *http://scaryscientist.blogspot.com/2015/10/outlier-detection.html*

Ramaswamy V, DeSarbo WS, Reibstein DJ, Robinson WT. An empirical pooling approach for estimating marketing mix elasticities with PIMS data. Marketing Science. 1993:103–124.

Reyna, C., & Brussino, S. (2011). Revisión de los fundamentos del análisis de clases latentes y ejemplo de aplicación en el área de las adicciones. *Trastornos adictivos*, *13*(1), 11-19.

Seo,S. (2002) A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets.

Tabachnick, B.G., Fidell, L.Sl. (2007): Using Multivariate Statistics, 5th ed. Allyn & Bacon/Pearson Education

Upton, G; Cook, I (1996). Understanding Statistics. Oxford University Press.

APPENDIX

Table A.1. The original survey questions selected

Media usage

1. First of all, we'd like to know how much time you spend on different media. Please think of yesterday:

How much time did you spend on the following media?

	Hours and minutes	Didn't use	Don't remember
Watched television on a TV set (flatscreen, etc.)	_____ Hours _______ Minutes	( )	( )
Watched television on a computer (PC, laptop, tablet, etc.)	_____ Hours _______ Minutes	( )	( )
Watched television on a mobile phone (iPhone, Nokia, HTC, etc.)	_____ Hours _______ Minutes	( )	( )
Listened to radio on a radio set (FM, DAB, etc.)	_____ Hours _______ Minutes	( )	( )
Listened to radio on a computer (PC, laptop, tablet, etc.)	_____ Hours _______ Minutes	( )	( )
Listened to radio on a mobile phone (iPhone, Nokia, HTC, etc.)	_____ Hours _______ Minutes	( )	( )
Read newspapers and magazines in the print version (on paper)	_____ Hours _______ Minutes	( )	( )
Read newspapers and magazines on the internet (at websites or designated applications)	_____ Hours _______ Minutes	( )	( )
Read books in the print version (on paper)	_____ Hours _______ Minutes	( )	( )
Read books in an electronic version [on a digital reader (Kindle, etc.), PC, laptop, tablet, mobile phone, etc.]	_____ Hours _______ Minutes	( )	( )
Listened to audio books	_____ Hours _______ Minutes	( )	( )

Different kinds of internet usage

15. Please think of yesterday – and any use you made of the internet yesterday.

How much time did you spend on the following things?

	Hours and minutes	Didn't use	Don't remember
Getting news (e.g., XXX, XXX)	_____ Hours _______ Minutes	( )	( )
Writing and reading e-mails	_____ Hours _______ Minutes	( )	( )
Downloading music, films, or podcasts	_____ Hours _______ Minutes	( )	( )
Playing computer games online	_____ Hours _______ Minutes	( )	( )
Using social network sites (e.g., Facebook, LinkedIn)	_____ Hours _______ Minutes	( )	( )
Using chat programs (e.g., Skype, WhatsApp)	_____ Hours _______ Minutes	( )	( )
Reading entries at debate sites, blogs, etc.	_____ Hours _______ Minutes	( )	( )
Writing entries at debate sites, blogs, etc. (including your own)	_____ Hours _______ Minutes	( )	( )
Online shopping, banking, travel reservations, etc.	_____ Hours _______ Minutes	( )	( )
Using websites concerning my interests or hobbies	_____ Hours _______ Minutes	( )	( )
Other – please specify	_____ Hours _______ Minutes	( )	( )

Please, skip questions 7-10 if you do not use a mobile phone.

7. Which functions do you use on your mobile phone? [checkbox]

❑ SMS (sending texts)
❑ MMS (sending images or sound)
❑ Watching TV or video (e.g., YouTube)
❑ Listening to radio
❑ Listening to podcasts
❑ Using phone as music player
❑ Taking photographs
❑ Recording video
❑ Visiting websites via browser
❑ Visiting websites via apps
❑ Instant messaging (e.g., WhatsApp)
❑ Social network sites (e.g., Facebook, LinkedIn)
❑ Games (e.g., Wordfeud, Angry Birds)
❑ Calendar
❑ Alarm clock and reminders
❑ E-mail
❑ GPS and maps
❑ Downloading apps
❑ Ordinary voice calls
❑ Other [Open]

Demographics

We have a few questions about yourself.

Gender

21. What is your sex? [radio, random]

❑ Male
❑ Female

Age

22. What is your age: [Open]

Education

23. How would you describe your family status?

❑ Single, no children
❑ Single, with children
❑ Married, no children
❑ Married, with children
❑ Divorced, no children
❑ Divorced, with children
❑ Widowed, no children
❑ Widowed, with children

24. Approximately how many years of education have you had? [radio, examples should be made by translator]

❑ 7 years or less
❑ About 8-9 years
❑ About 10-11 years (e.g., vocational training)
❑ About 12 years (e.g., high school)
❑ About 13-14 years (e.g., technical education)
❑ About 15 years (e.g., Bachelor's degree)
❑ About 16-17 years (e.g., Master's degree)
❑ 18 years or more (e.g., PhD)
❑ Don't know

Income

25. The average monthly personal income in [country] is [.......] before taxes. What is your monthly income?

❑ A lot above average
❑ Slightly above average
❑ Similar to the average
❑ Slightly below average
❑ A lot below average
❑ Don't know
❑ Prefer not to respond

Employment

26. What is your employment status? [checkbox, random]

❑ Full-time work
❑ Part-time work
❑ Unemployed
❑ Retired
❑ In unpaid position (housework, volunteer or community service, military service, etc.)
❑ Other [open]
❑ Don't know

Table A.2. Description of the original database variables

SOCIODEMOGRAPHIC VARIABLES	
Q21_Sex	Gender
Q22_Age	Age
Q23_Family_status	Family status
Q24_Education	Education level
Q25_Income	Monthly income
Q26_Employment	Employment status
MEDIA & INTERNET USAGE	
Q1_TV_TVSET_HOUR	Watched television on a tv set / Hours
Q1_TV_TVSET_MIN	Watched television on a tv set / Minutes
Q1_TV_TVSET_NO	Watched television on a tv set / Did not use
Q1_TV_TVSET_REM	Watched television on a tv set / Do not remember
Q1_TV_COMP_HOUR	Watched television on a computer / Hours
Q1_TV_COMP_MIN	Watched television on a computer / Minutes
Q1_TV_COMP_NO	Watched television on a computer/ Did not use
Q1_TV_COMP_REM	Watched television on a computer / Do not remember
Q1_TV_MOB_HOUR	Watched television on a mobile phone / Hours
Q1_TV_MOB_MIN	Watched television on a mobile phone / Minutes
Q1_TV_MOB_NO	Watched television on a mobile phone / Did not use
Q1_TV_MOB_REM	Watched television on a mobile phone / Do not remember
Q1_RAD_SET_HOUR	Listened to radio on a radio set / Hours
Q1_RAD_SET_MIN	Listened to radio on a radio set / Minutes
Q1_RAD_SET_NO	Listened to radio on a radio set / Did not use
Q1_RAD_SET_REM	Listened to radio on a radio set / Do not remember
Q1_RAD_COMP_HOUR	Listened to radio on computer / Hours
Q1_RAD_COMP_MIN	Listened to radio on computer / Minutes
Q1_RAD_COMP_NO	Listened to radio on a computer/ Did not use
Q1_RAD_COMP_REM	Listened to radio on a computer / Do not remember
Q1_RAD_MOB_HOUR	Listened to radio on mobile phone / Hours
Q1_RAD_MOB_MIN	Listened to radio on mobile phone / Minutes
Q1_RAD_MOB_NO	Listened to radio on a mobile phone / Did not use
Q1_RAD_MOB_REM	Listened to radio on a mobile phone / Do not remember

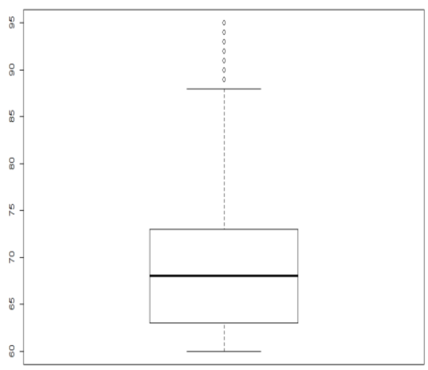
Q1_NEWSP_PRINT_HOUR	Read newspapers or magazines in the printed version / Hours
Q1_NEWSP_PRINT_MIN	Read newspapers or magazines in the printed version / Minutes
Q1_NEWSP_PRINT_NO	Read newspapers or magazines in the printed version / Did not use
Q1_NEWSP_PRINT_REM	Read newspapers or magazines in the printed version / Do not remember
Q1_NEWSP_INT_HOUR	Read newspapers or magazines on the internet / Hours
Q1_NEWSP_INT_MIN	Read newspapers or magazines on the internet / Minutes
Q1_NEWSP_INT_NO	Read newspapers or magazines on the internet / Did not use
Q1_NEWSP_INT_REM	Read newspapers or magazines on the internet / Do not remember
Q1_BOOKS_PRINT_HOUR	Read books in the printed version / Hours
Q1_BOOKS_PRINT_MIN	Read books in the printed version / Minutes
Q1_BOOKS_PRINT_NO	Read books in the printed version / Did not use
Q1_BOOKS_PRINT_REM	Read books in the printed version / Do not remember
Q1_BOOKS_ELEC_HOUR	Read books in the electronic version / Hours
Q1_BOOKS_ELEC_MIN	Read books in the electronic version / Minutes
Q1_BOOKS_ELEC_NO	Read books in the electronic version / Did not use
Q1_BOOKS_ELEC_REM	Read books in the electronic version / Do not remember
Q1_AUDBOOKS_HOUR	Listened to audio books : Hours
Q1_AUDBOOKS_MIN	Listened to audio books : Minutes
Q1_AUDBOOKS_NO	Listened to audio books / Did not use
Q1_AUDBOOKS_REM	Listened to audio books / Do not remember
Q15_Inter_news_HOUR	Internet use yesterday / Getting news / Hours
Q15_Inter_news_MIN	Internet use yesterday / Getting news / Minutes
Q15_Inter_news_NO	Internet use yesterday / Getting news / Did not use
Q15_Inter_news_REM	Internet use yesterday / Getting news / Do not remember
Q15_Inter_emails_HOUR	Internet use yesterday / Wrinting and reading e-mails / Hours
Q15_Inter_emails_MIN	Internet use yesterday / Writing and reading e-mails / Minutes
Q15_Inter_emails_NO	Internet use yesterday / Writing and reading e-mails / Did not use
Q15_Inter_emails_REM	Internet use yesterday / Writing and reading e-mails / Do not remember
Q15_Inter_podcast_HOUR	Internet use yesterday / Downloading music, film or podcasts / Hours
Q15_Inter_podcast_MIN	Internet use yesterday / Downloading music, film or podcasts / Minutes
Q15_Inter_podcast_NO	Internet use yesterday / Downloading music, film or podcasts / Did not use
Q15_Inter_podcast_REM	Internet use yesterday / Downloading music, film or podcasts / Do not remember
Q15_Inter_games_HOUR	Internet use yesterday / Playing computer games online / Hours
Q15_Inter_games_MIN	Internet use yesterday / Playing computer games online / Minutes
Q15_Inter_games_NO	Internet use yesterday / Playing computer games online / Did not use
Q15_Inter_games_REM	Internet use yesterday / Playing computer games online / Do not remember

Q15_Inter_SNS_HOUR	Internet use yesterday / Using social network sites / Hours
Q15_Inter_SNS_MIN	Internet use yesterday / Using social network sites / Minutes
Q15_Inter_SNS_NO	Internet use yesterday / Using social network sites / Did not use
Q15_Inter_SNS_REM	Internet use yesterday / Using social network sites / Do not remember
Q15_Inter_chat_HOUR	Internet use yesterday / Using chat programs / Hours
Q15_Inter_chat_MIN	Internet use yesterday / Using chat programs / Minutes
Q15_Inter_chat_NO	Internet use yesterday / Using chat programs / Did not use
Q15_Inter_chat_REM	Internet use yesterday / Using chat programs / Do not remember
Q15_Inter_readblogs_HOUR	Internet use yesterday / Reading entries at debate sites, blogs / Hours
Q15_Inter_readblogs_MIN	Internet use yesterday / Reading entries at debate sites, blogs / Minutes
Q15_Inter_readblogs_NO	Internet use yesterday / Reading entries at debate sites, blogs / Did not use
Q15_Inter_readblogs_REM	Internet use yesterday / Reading entries at debate sites, blogs / Do not remember
Q15_Inter_writeblogs_HOUR	Internet use yesterday / Writing entries at debate sites, blogs / Hours
Q15_Inter_writeblogs_MIN	Internet use yesterday / Writing entries at debate sites, blogs / Minutes
Q15_Inter_writeblogs_NO	Internet use yesterday / Writing entries at debate sites, blogs / Did not use
Q15_Inter_writeblogs_REM	Internet use yesterday / Writing entries at debate sites, blogs / Do not remember
Q15_Inter_shopping_HOUR	Internet use yesterday / Online shopping, banking, travel reservation etc. / Hours
Q15_Inter_shopping_MIN	Internet use yesterday / Online shopping, banking, travel reservation etc. / Minutes
Q15_Inter_shopping_NO	Internet use yesterday / Online shopping, banking, travel reservation etc. / Did not use
Q15_Inter_shopping_REM	Internet use yesterday / Online shopping, banking, travel reservation etc. / Do not remember
Q15_Inter_hobbies_HOUR	Internet use yesterday / Using websites concerning my interests or hobbies / Hours
Q15_Inter_hobbies_MIN	Internet use yesterday / Using websites concerning my interests or hobbies / Minutes
Q15_Inter_hobbies_NO	Internet use yesterday / Using websites concerning my interests or hobbies / Did not use
Q15_Inter_hobbies_REM	Internet use yesterday / Using websites concerning my interests or hobbies / Do not remember
Q15_Inter_other_HOUR	Internet use yesterday / Other / Hours
Q15_Inter_other_MIN	Internet use yesterday / Other / Minutes
Q15_Inter_other_NO	Internet use yesterday / Other / Did not use
Q15_Inter_other_REM	Internet use yesterday / Other / Do not remember
MOBILE PHONE USAGE	
Q7_SMS	Mobile Phone - SMS
Q7_MMS	Mobile Phone -MMS (Multimedia Message Services)

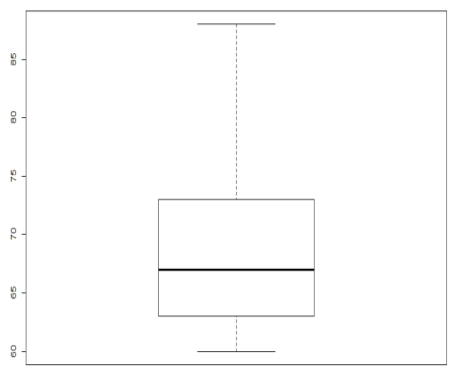
Q7_TV_VIDEO	Mobile Phone - Watching TV or video on mobile
Q7_RADIO	Mobile Phone - Listening to radio
Q7_PODCAST	Mobile Phone - Listening to podcast
Q7_MUSIC_PLAYER	Mobile Phone - Using a phone as a music player
Q7_PHOTOS	Mobile Phone - Taking photographs
Q7_REC_VIDEO	Mobile Phone - Recording video
Q7_WEB_BROWSER	Mobile Phone - Viewing websites via browser
Q7_WEB_APPS	Mobile Phone -Viewing websites via apps
Q7_INST_MESS	Mobile Phone - Instant messaging
Q7_SNS	Mobile Phone - Social network sites
Q7_GAMES	Mobile Phone - Games
Q7_CALENDAR	Mobile Phone - Calendar
Q7_ALARM	Mobile Phone - Alarm clock and reminders
Q7_EMAIL	Mobile Phone - E-mail
Q7_GPS_MAPPS	Mobile Phone - GPS and maps
Q7_DOWN_APPS	Mobile Phone - Downloading apps
Q7_VOICE_CALLS	Mobile Phone - Ordinary voice calls
Q7_OTHER	Mobile Phone -Other (mobile usage)
OTHERS	
ID	ID merged dataset
COUNTRYCODE	Country code
POND	Ponderation recoding only for Canada

Figure A.1 Boxplot comparison of Age before remove the outliers and after by Z-score method

Age variable of Canada



Before



After

Age variable of Spain

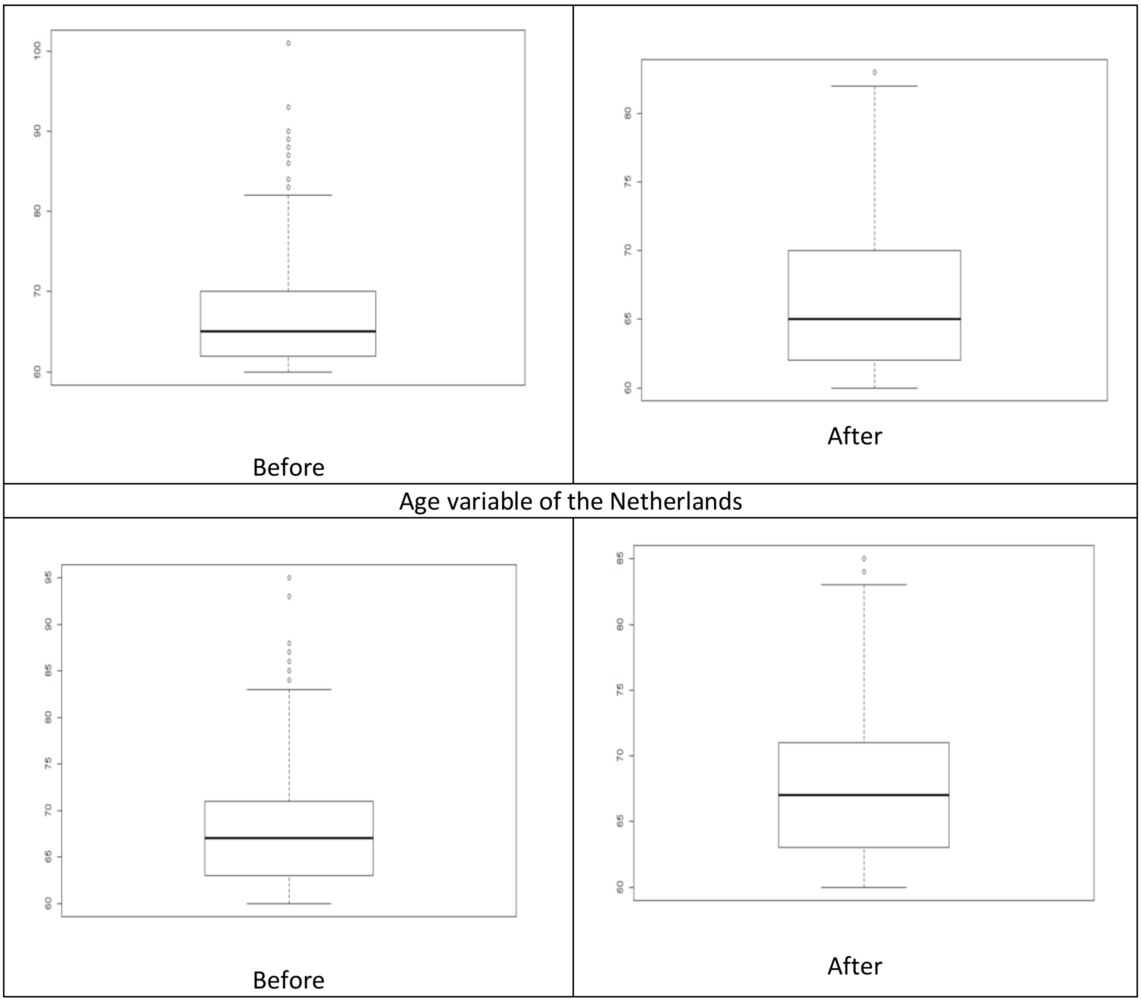
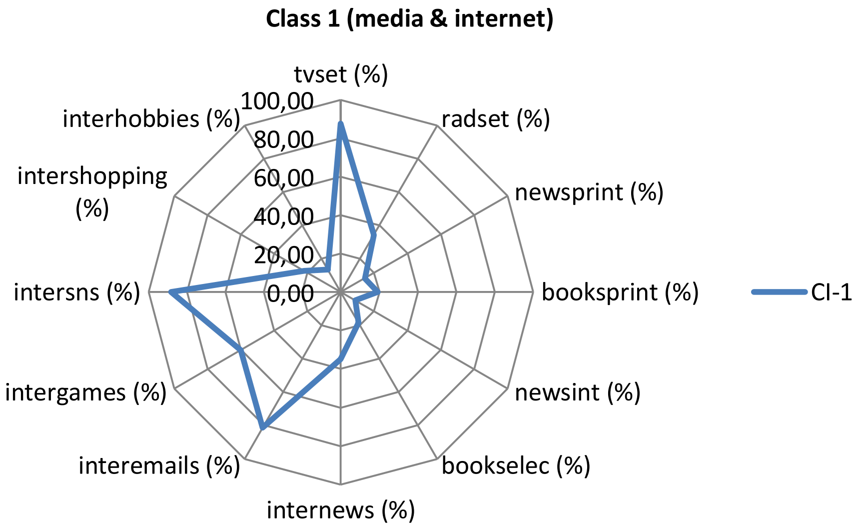


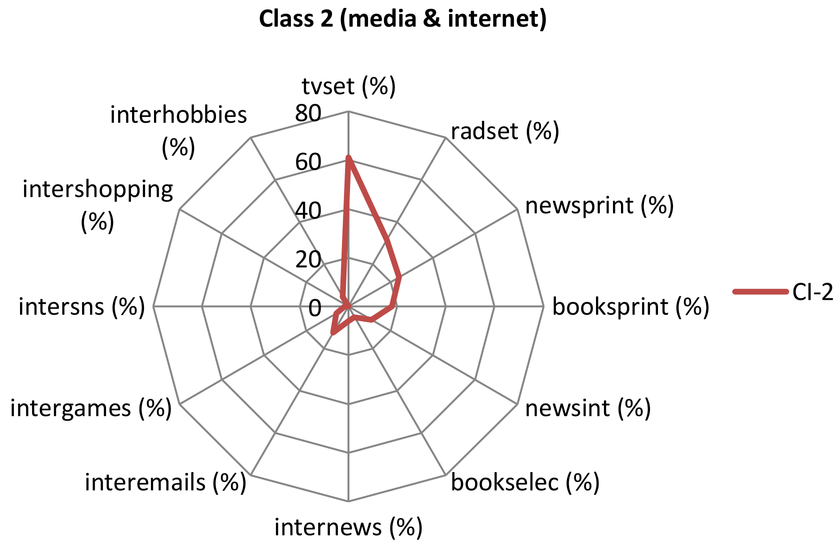
Figure A.2. Shapiro normality test for the age variable by country

```
> shapiro.test(canada$Q22_Age)

        Shapiro-Wilk normality test

data:  canada$Q22_Age
W = 0.9382, p-value < 2.2e-16

> shapiro.test(spain$Q22_Age)

        Shapiro-Wilk normality test

data:  spain$Q22_Age
W = 0.905, p-value < 2.2e-16

> shapiro.test(holland$Q22_Age)

        Shapiro-Wilk normality test

data:  holland$Q22_Age
W = 0.9325, p-value < 2.2e-16
```

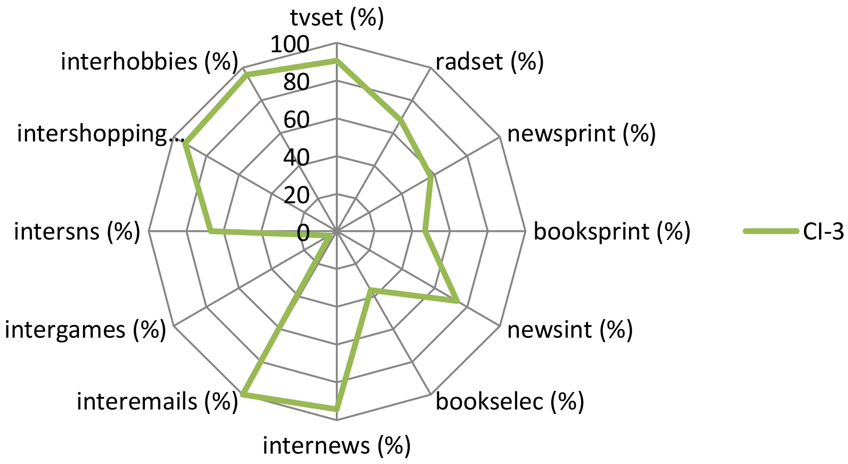
Table A.3. Results of media & internet cluster (Canada)

Class	N	Percentage	tvset (%)	radset (%)	newsprint (%)	booksprint (%)	newsint (%)	bookselec (%)	internews (%)	interemails (%)	intergames (%)	intersns (%)	intershopping (%)	interhobbies (%)
Cl-1	524	14,88%	87,40	34,16	14,30	19,27	8,59	18,51	34,92	81,49	60,11	88,17	22,14	13,36
Cl-2	321	9,11%	61,06	31,46	23,99	17,76	10,90	4,98	5,92	12,46	5,61	1,56	0,00	4,67
Cl-3	549	15,59%	90,35	67,94	57,74	46,81	73,77	36,07	94,17	99,82	4,53	66,85	93,08	95,81
Cl-4	945	26,83%	84,87	59,47	46,14	44,55	100,00	15,45	99,58	89,84	35,66	57,99	24,87	28,25
Cl-5	1183	33,59%	94,67	66,10	78,19	56,80	5,41	6,59	38,04	90,19	32,97	41,25	28,15	31,11
TOTAL	3522	100%	87,22	56,70	51,96	42,81	42,42	15,19	59,91	83,22	36,90	53,10	33,93	35,38

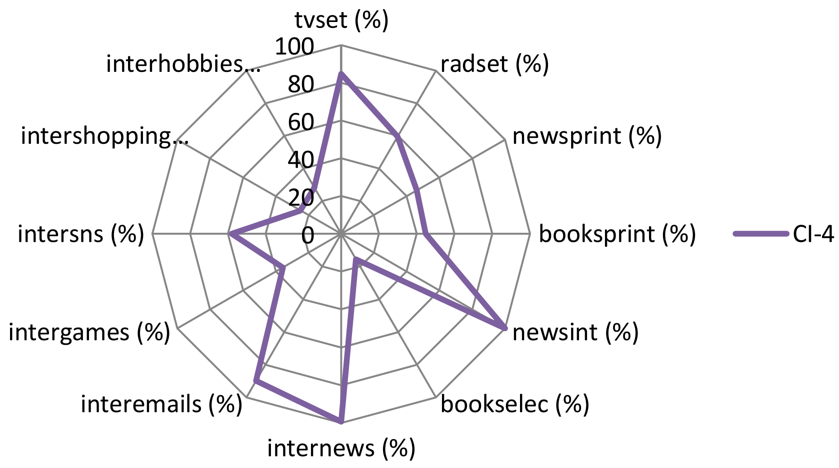




Class 3 (media & internet)



Class 4 (media & internet)



Class 5 (media & internet)

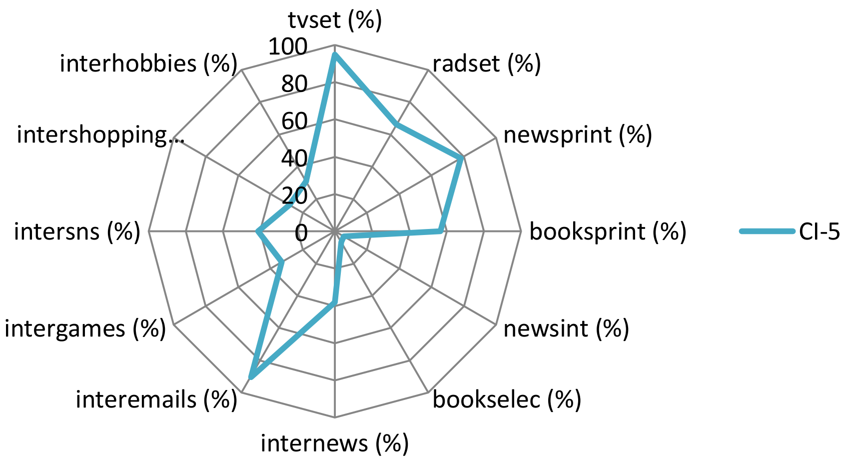


Table A.4. Canada chi-square test results (α=0.05)

CANADA	CI	CM
Gender	p = 4.697443e-12 Minimum expected frequency: 150.5674	p = 3.896075e-07 Minimum expected frequency: 181.2828
Age 5 categories	p = 5.031074e-07 Minimum expected frequency: 29.56698	p = 1.329335e-50 Minimum expected frequency: 30.52534
Has partner	p = 0.01869985 Minimum expected frequency: 115.6757	p = 0.05792616 Minimum expected frequency: 125.6073
Has children	no p = 0.1826147 Minimum expected frequency: 151.476	no p = 0.1737111 Minimum expected frequency: 184.1734
Educational level	p = 6.693636e-36 Minimum expected frequency: 7.88995	p = 1.151779e-06 Minimum expected frequency: 7.631603
Income	p = 1.984047e-17 Minimum expected frequency: 36.34778	p = 0.0001681645 Minimum expected frequency: 43.056
Employment	p=0,04233095 Minimum expected frequency: 53.21587	p= 4.741642e-19 Minimum expected frequency: 66.04123
CM	p = 1.832107e-29 Minimum expected frequency: 33.65468	-

***CI= Clusters of media & internet, CM=Clusters of mobile phone usage (same for the following tables)**

Table A.5. Cross-table outputs of Canada media & internet usage

Gender (N=3515)	**CI-1**	**CI-2**	**CI-3**	**CI-4**	**CI-5**	**Column Total**
Male	32.3%	42.5%	49.8%	52.3%	46.9%	46.2%
Female	67.7%	57.5%	50.2%	47.7%	53.1%	53.8%
TOTAL	100%	100%	100%	100%	100%	100%

Age in 5 categories (N=3515)	**CI-1**	**CI-2**	**CI-3**	**CI-4**	**CI-5**	**Column Total**
[60,65[	29.8%	27.2%	33.1%	30.0%	23.5%	27.9%
[65,70[	22.9%	24.4%	26.4%	25.1%	23.0%	24.2%
[70,75[	19.6%	18.0%	14.9%	18.9%	16.2%	17.4%
[75,80[	21.7%	22.1%	18.4%	17.2%	25.7%	21.4%
[80,+ [	5.9%	8.4%	7.3%	8.8%	11.6%	9.1%

TOTAL	100%	100%	100%	100%	100%	100%

Has partner (N=3515)	CI-1	CI-2	CI-3	CI-4	CI-5	Column total
Married	58.8%	63.4%	67.2%	67.0%	64.2%	64.5%
Not married	41.2%	36.6%	32.8%	33.0%	35.8%	35.5%
TOTAL	100%	100%	100%	100%	100%	100%

Education (N=3505)	CI-1	CI-2	CI-3	CI-4	CI-5	Column total
Primary or less	2.8%	7.6%	0.9%	1.7%	2.1%	2.4%
Secondary	73.1%	67.5%	48.3%	47.7%	56.1%	56.3%
Tertiary	24.1%	24.9%	50.9%	50.6%	41.8%	41.3%
TOTAL	100%	100%	100%	100%	100%	100%

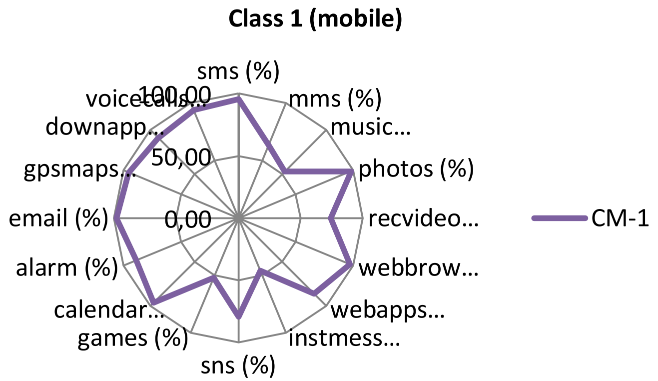
Income in 3 categories (N=3515)	CI-1	CI-2	CI-3	CI-4	CI-5	Column total
Above the average	31.9%	32.4%	55.4%	49.9%	43.3%	44.1%
Similar to average	12.8%	10.8%	10.9%	10.7%	11.0%	11.2%
Below the average	42.8%	37.7%	24.6%	27.6%	31.4%	31.7%
Not declared	12.5%	19.1%	9.2%	11.8%	14.4%	13.1%
TOTAL	100%	100%	100%	100%	100%	100%

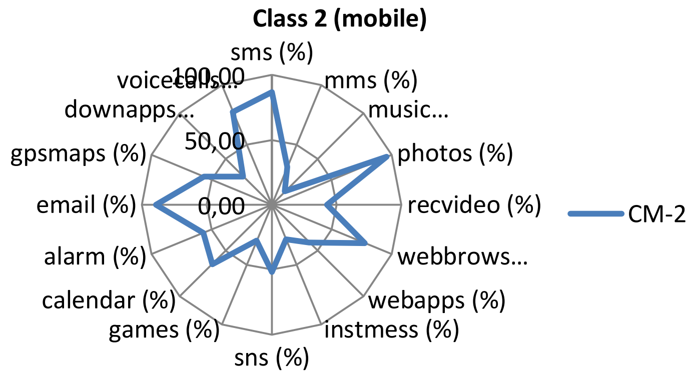
Employ (N=3394)	CI-1	CI-2	CI-3	CI-4	CI-5	Column total
Active	14.7%	14.6%	21.1%	17.2%	16.2%	16.8%
Inactive	85.3%	85.4%	78.9%	82.8%	83.8%	83.2%
TOTAL	100%	100%	100%	100%	100%	100%

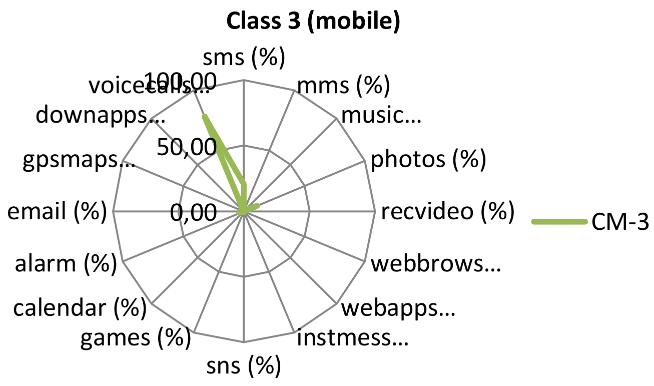
Table A.6. Results of mobile phone features cluster (Canada)

Class	N	Percentage	sms (%)	mms (%)	music player (%)	photos (%)	recvideo (%)	webbrowser (%)	webapps (%)	instmess (%)	sns (%)	games (%)	calendar (%)	alarm (%)	email (%)	gpsmaps (%)	downapps (%)	voicecalls (%)
CM-1	443	15,24%	95,70	63,40	52,80	98,20	74,30	97,30	86,20	45,80	79,50	51,90	96,40	88,90	98,20	95,70	91,00	93,70
CM-2	756	26,01%	86,60	31,70	14,60	96,00	42,60	77,90	40,90	29,10	52,10	29,90	64,90	57,00	88,90	56,30	30,80	77,60
CM-3	869	29,89%	20,90	0,00	0,70	10,80	0,00	0,10	0,00	0,60	0,00	0,20	1,20	3,70	3,10	0,20	0,00	78,00
CM-4	839	28,86%	77,40	12,30	6,10	83,70	14,40	12,50	4,80	10,40	9,50	11,40	37,80	36,60	35,00	10,70	3,20	69,60
TOTAL	2464	100%	65,69	21,46	13,81	67,32	26,56	38,73	25,16	17,73	28,41	19,04	42,84	40,04	49,12	32,38	22,80	77,86

canada






Class 3 (mobile)
sms (%)
mms (%)
music...
photos (%)
recvideo (%)
webbrows...
webapps...
instmess...
sns (%)
games (%)
calendar (%)
alarm (%)
email (%)
gpsmaps...
downapps...
voicecalls...
100,00
50,00
0,00
CM-3

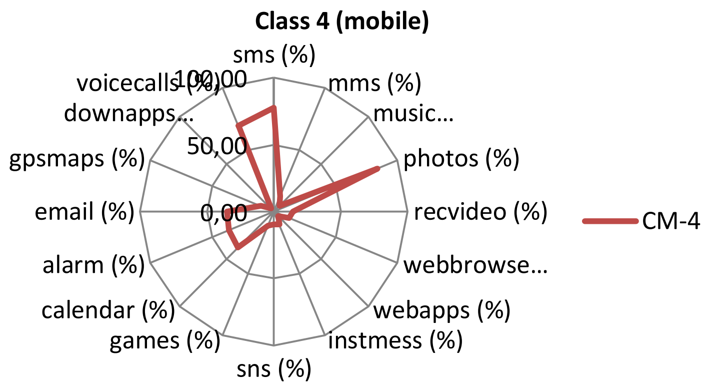

Class 4 (mobile)
sms (%)
mms (%)
music...
photos (%)
recvideo (%)
webbrowse...
webapps (%)
instmess (%)
sns (%)
games (%)
calendar (%)
alarm (%)
email (%)
gpsmaps (%)
downapps...
voicecalls (%)
100,00
50,00
0,00
CM-4

Table A.7. Cross-table outputs of Canada mobile phone usage

Gender (N=2871)	**CM-1**	**CM-2**	**CM-3**	**CM-4**	**Column total**
Male	55.5%	51.1%	40.5%	45.5%	46.6%
Female	44.5%	48.9%	59.5%	54.5%	53.4%
TOTAL	100%	100%	100%	100%	100%

Age (N=2871)	**CM-1**	**CM-2**	**CM-3**	**CM-4**	**Column Total**
[60,65[	47.5%	33.8%	18.4%	27.1%	28.6%
[65,70[	26.1%	28.8%	19.6%	25.4%	24.4%
[70,75[	14.6%	17.7%	16.7%	21.2%	18.0%
[75,80[	9.0%	15.9%	31.8%	19.1%	21.2%
[80,+ [	2.9%	3.8%	13.5%	7.1%	7.8%
TOTAL	100%	100%	100%	100%	100%

Education (N=2863)	**CM-1**	**CM-2**	**CM-3**	**CM-4**	**Column total**
Primary or less	0.7%	1.6%	2.3%	2.5%	2.0%
Secondary	47.9%	52.3%	62.1%	54.8%	55.7%
Tertiary	51.4%	46.1%	35.6%	42.7%	42.4%
TOTAL	100%	100%	100%	100%	100%

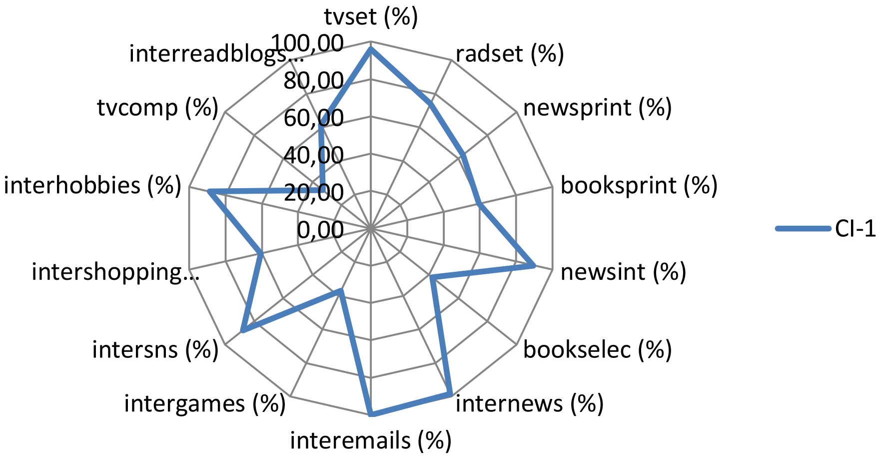
Income (N=2871)	**CM-1**	**CM-2**	**CM-3**	**CM-4**	**Column total**
Above the average	55.2%	50.9%	43.0%	43.4%	46.7%
Similar to average	8.6%	9.8%	12.1%	12.2%	11.1%
Below the average	24.7%	26.2%	33.6%	30.6%	29.7%
Not declared	11.5%	13.1%	11.3%	13.8%	12.5%
TOTAL	100%	100%	100%	100%	100%

Employment (N=2764)	**CM-1**	**CM-2**	**CM-3**	**CM-4**	**Column total**
Active	29.4%	23.5%	9.8%	18.1%	18.1%
Inactive	70.6%	76.5%	90.2%	81.9%	81.9%
TOTAL	100%	100%	100%	100%	100%

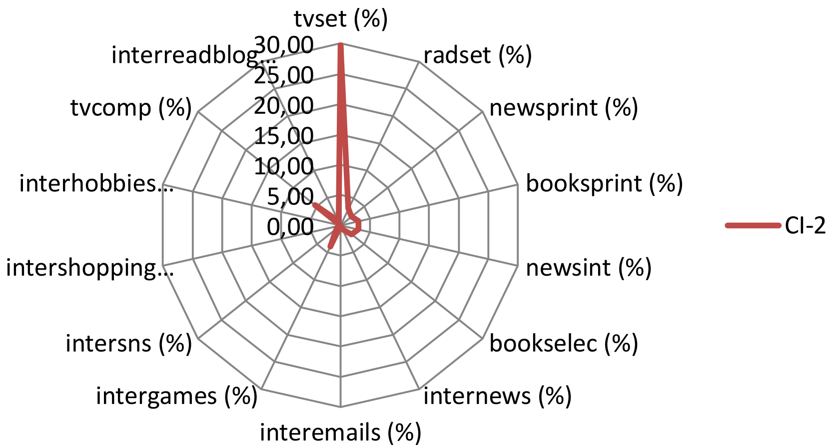
Table.A.8. Results of media & internet cluster (Spain)

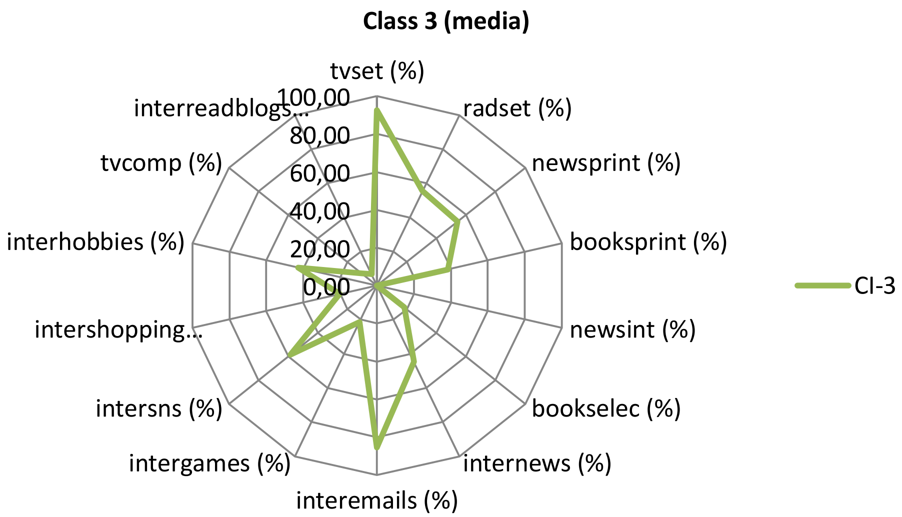
Class	N	Percentage	tvset (%)	radset (%)	newsprint (%)	booksprint (%)	newsint (%)	bookselec (%)	internews (%)	interemails (%)	intergames (%)	intersns (%)	intershopping (%)	interhobbies (%)	tvcomp (%)	interreadblogs (%)	radcomp (%)	interchat (%)
Cl-1	323	14,52%	95,98	73,99	63,16	59,44	89,47	42,11	98,45	100,00	37,15	87,62	60,68	88,85	32,82	61,61	32,20	92,88
Cl-2	129	5,80%	29,68	3,10	2,33	3,10	3,10	2,33	0,78	0,00	3,88	0,78	0,00	0,00	5,43	0,78	4,65	1,55
Cl-3	597	26,84%	92,80	55,44	54,44	38,36	0,00	18,43	44,72	85,43	21,44	58,29	20,10	42,88	10,22	6,70	8,21	56,11
Cl-4	297	13,35%	98,32	62,96	50,17	41,75	23,23	14,81	3,37	20,54	1,68	6,40	0,34	2,69	8,42	0,34	2,69	16,16
Cl-5	878	39,48%	96,24	64,81	51,03	43,17	100,00	27,90	86,22	83,37	17,88	55,81	21,87	46,70	13,33	10,71	9,79	51,48
TOTAL	**2224**	100%	91,69	59,80	50,76	41,72	55,75	24,19	60,83	73,11	18,66	51,30	22,89	43,21	14,21	15,06	11,37	51,12

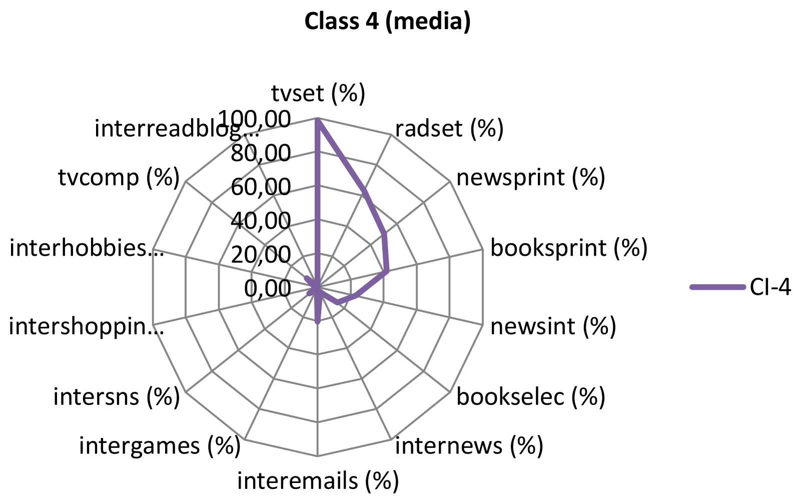
Class 1 (media)



Class 2 (media)




Class 3 (media)
tvset (%)
radset (%)
newsprint (%)
booksprint (%)
newsint (%)
bookselec (%)
internews (%)
interemails (%)
intergames (%)
intersns (%)
intershopping...
interhobbies (%)
tvcomp (%)
interreadblogs...
100,00
80,00
60,00
40,00
20,00
0,00
Cl-3


Class 4 (media)
tvset (%)
radset (%)
newsprint (%)
booksprint (%)
newsint (%)
bookselec (%)
internews (%)
interemails (%)
intergames (%)
intersns (%)
intershoppin...
interhobbies...
tvcomp (%)
interreadblog...
100,00
80,00
60,00
40,00
20,00
0,00
Cl-4

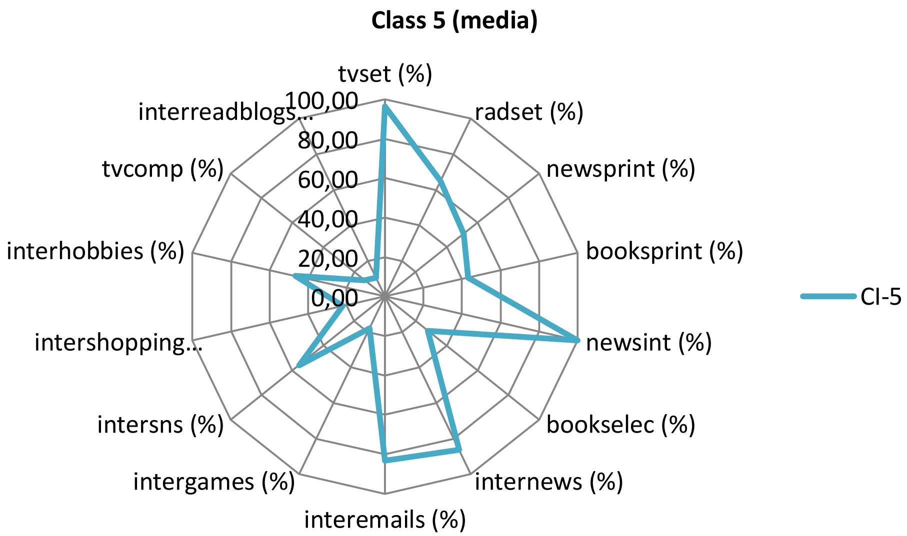

Class 5 (media)
tvset (%)
radset (%)
newsprint (%)
booksprint (%)
newsint (%)
bookselec (%)
internews (%)
interemails (%)
intergames (%)
intersns (%)
intershopping...
interhobbies (%)
tvcomp (%)
interreadblogs...
100,00
80,00
60,00
40,00
20,00
0,00
Cl-5

Table A.9. Spain chi-square test results (α=0.05)

SPAIN	CI	CM
Gender	p = 4.965706e-13 Minimum expected frequency: 59.6277	p = 0.3223409 Minimum expected frequency: 74.34826
Age 5 categories	p = 6.227019e-09 Minimum expected frequency:12.23876	p=5.86965e-27 Minimum expected frequency: 15.60199
Has partner	p = 0.2764969 Minimum expected frequency: 28.29358	p = 0.5963378 Minimum expected frequency: 37.28203
Has children	p = 0.3704486 Minimum expected frequency: 51.02752	p = 0.9118864 Minimum expected frequency: 65.08557
Educational level	p = 1.074899e-22 Minimum expected frequency: 26.65077	p = 5.652405e-15 Minimum expected frequency: 33.48622
Income	p = 3.086085e-23 Minimum expected frequency: 14.38489	p-value = 8.459e-09 Minimum expected frequency: 17.75124
Employment	p = 0.01740203 Minimum expected frequency: 28.04872	p = 2.094468e-09 Minimum expected frequency: 34.06751
CM	p = 4.381409e-79 Minimum expected frequency: 8.199005	-

Table A.10. Cross-table outputs of Spain media & internet usage

Gender (N=2224)	**CI-1**	**CI-2**	**CI-3**	**CI-4**	**CI-5**	**Column total**
Male	55.7%	37.2%	49.4%	41.8%	62.5%	53.8%
Female	44.3%	62.8%	50.6%	58.2%	37.5%	46.2%
TOTAL	100%	100%	100%	100%	100%	100%

Age in 5 categories (N=2224)	**CI-1**	**CI-2**	**CI-3**	**CI-4**	**CI-5**	**Column Total**
[60,65[	55.4%	47.3%	47.7%	41.4%	48.5%	48.3%
[65,70[	22.0%	26.4%	24.0%	18.2%	21.5%	22.1%
[70,75[	18.3%	16.3%	19.6%	20.5%	21.6%	20.1%
[75, +]	4.3%	10.1%	8.7%	19.9%	8.3%	9.5%
TOTAL	100%	100%	100%	100%	100%	100%

* Categories [75,80[and [80,+[are merged to avoid cells with frequencies below 5.

Has partner (N=2180)	**CI-1**	**CI-2**	**CI-3**	**CI-4**	**CI-5**	**Column total**
Married	74.1%	75.0%	75.5%	74.0%	78.9%	76.4%
Not married	25.9%	25.0%	24.5%	26.0%	21.1%	23.6%
TOTAL	100%	100%	100%	100%	100%	100%

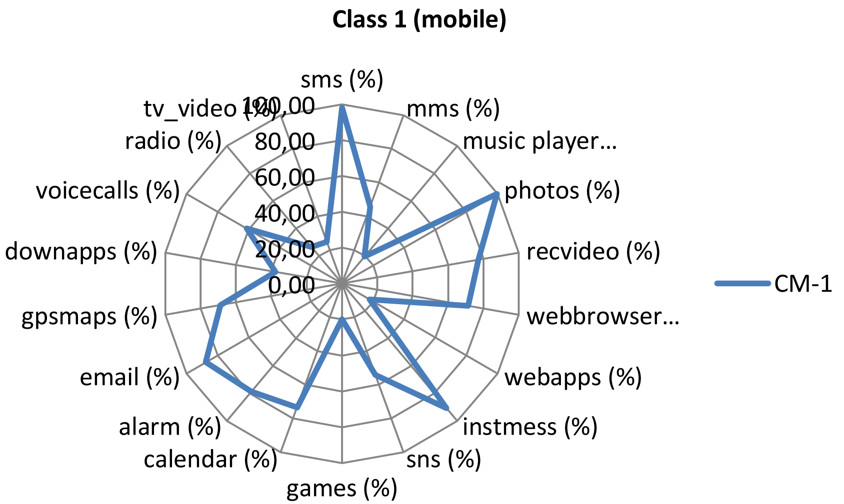
Has children (N=2180)	**CI-1**	**CI-2**	**CI-3**	**CI-4**	**CI-5**	**Column total**
Yes	59.2%	62.5%	55.7%	54.0%	58.6%	57.5%
No	40.8%	37.5%	44.3%	46.0%	41.4%	42.5%
TOTAL	100%	100%	100%	100%	100%	100%
Education (N=2202)	CI-1	CI-2	CI-3	CI-4	CI-5	Column total
Primary or less	11.8%	35.5%	24.1%	39.2%	16.8%	22.0%
Secondary	34.1%	38.0%	41.4%	32.3%	38.4%	37.7%
Tertiary	54.1%	26.5%	34.5%	28.5%	44.8%	40.3%
TOTAL	100%	100%	100%	100%	100%	100%

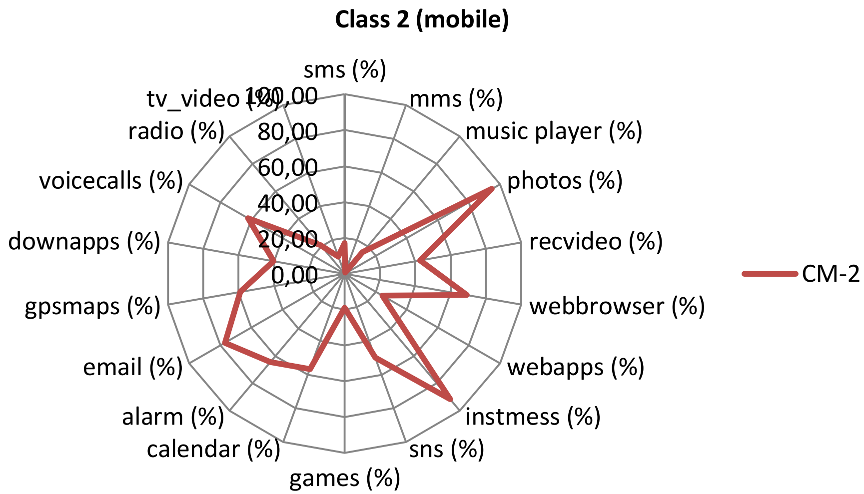
Income (N=2224)	**CI-1**	**CI-2**	**CI-3**	**CI-4**	**CI-5**	**Column total**
Above the average	58.1%	17.1%	48.2%	28.6%	52.6%	47.8%
Similar to average	14.2%	26.4%	17.8%	22.9%	14.4%	17.1%
Below the average	11.2%	14.0%	11.1%	18.5%	8.3%	11.2%
Not declared	16.1%	42.6%	23.0%	30.0%	22.8%	24.0%
TOTAL	100%	100%	100%	100%	100%	100%

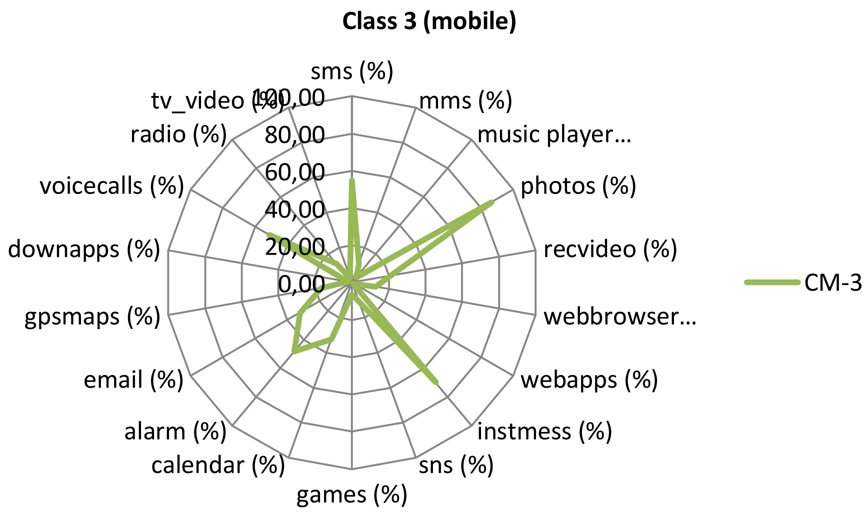
Employment (N=2218)	**CI-1**	**CI-2**	**CI-3**	**CI-4**	**CI-5**	**Column total**
Active	28.2%	23.6%	19.0%	19.0%	22.8%	22.1%
Inactive	71.8%	76.4%	81.0%	81.0%	77.2%	77.9%
TOTAL	100%	100%	100%	100%	100%	100%

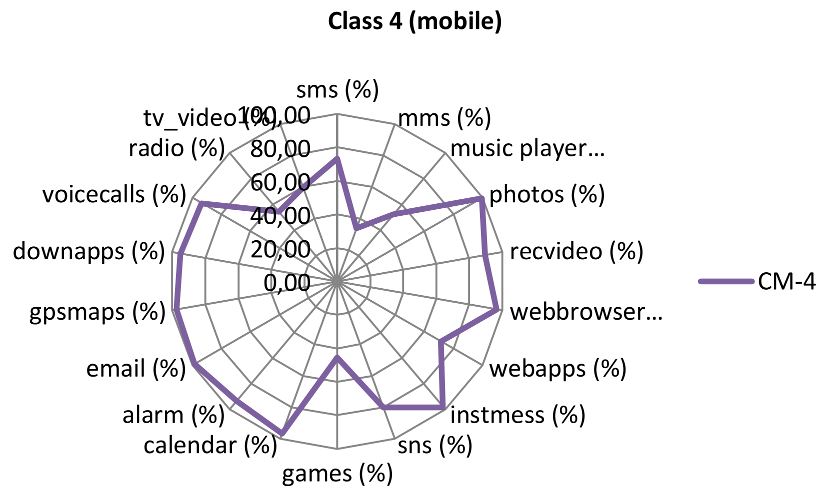
Table.A.11. Results of mobile phone features cluster (Spain)

Class	N	Percentage	sms (%)	mms (%)	music player (%)	photos (%)	recvideo (%)	webbrowser (%)	webapps (%)	instmess (%)	sns (%)	games (%)	calendar (%)	alarm (%)	email (%)	gpsmaps (%)	downapps (%)	voicecalls (%)	radio (%)	tv_video (%)
CM-1	417	20,75%	98,30	45,60	19,70	99,80	77,00	71,50	17,70	90,60	54,00	19,90	73,40	78,70	87,80	68,80	37,60	61,20	26,60	24,70
CM-2	426	21,19%	17,10	0,50	15,50	94,80	42,70	69,20	24,40	91,30	50,00	19,00	56,80	64,60	77,20	58,90	40,10	62,20	20,90	10,30
CM-3	575	28,61%	54,60	10,80	3,80	86,40	20,90	12,70	0,90	69,90	10,80	6,40	32,20	48,20	32,30	16,90	2,30	51,30	13,00	2,80
CM-4	432	21,49%	73,40	33,60	52,30	99,50	89,60	96,50	71,50	98,10	80,30	45,40	96,80	93,30	98,60	97,50	95,10	93,50	54,20	59,70
CM-5	160	7,96%	34,40	0,00	0,60	11,90	0,00	0,00	0,00	11,90	0,00	0,00	0,00	11,90	0,60	0,60	0,00	86,90	2,50	0,60
TOTAL	**2010**	**100%**	**58,15**	**19,88**	**19,75**	**87,85**	**50,26**	**53,87**	**24,47**	**80,17**	**42,15**	**19,74**	**57,28**	**64,81**	**65,06**	**52,59**	**37,39**	**67,57**	**25,51**	**20,99**






Class 3 (mobile)
sms (%)
mms (%)
music player...
photos (%)
recvideo (%)
webbrowser...
webapps (%)
instmess (%)
sns (%)
games (%)
calendar (%)
alarm (%)
email (%)
gpsmaps (%)
downapps (%)
voicecalls (%)
radio (%)
tv_video (%)
100,00
80,00
60,00
40,00
20,00
0,00
CM-3


Class 4 (mobile)
sms (%)
mms (%)
music player...
photos (%)
recvideo (%)
webbrowser...
webapps (%)
instmess (%)
sns (%)
games (%)
calendar (%)
alarm (%)
email (%)
gpsmaps (%)
downapps (%)
voicecalls (%)
radio (%)
tv_video (%)
100,00
80,00
60,00
40,00
20,00
0,00
CM-4

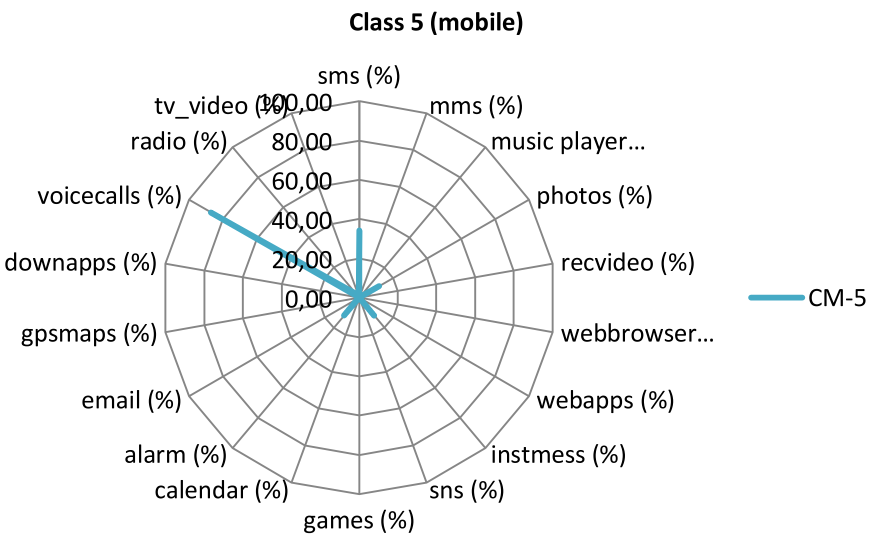

Class 5 (mobile)
sms (%)
mms (%)
music player...
photos (%)
recvideo (%)
webbrowser...
webapps (%)
instmess (%)
sns (%)
games (%)
calendar (%)
alarm (%)
email (%)
gpsmaps (%)
downapps (%)
voicecalls (%)
radio (%)
tv_video (%)
100,00
80,00
60,00
40,00
20,00
0,00
CM-5

Table A.12. Cross-table outputs of Spain mobile phone usage

Age in 5 categories (N=2010)	**CM-1**	**CM-2**	**CM-3**	**CM-4**	**CM-5**	**Column total**
[60,65[	46.3%	55.0%	42.0%	57.9%	23.8%	47.5%
[65,70[	24.0%	22.5%	20.5%	23.6%	22.5%	22.5%
[70,75[	22.0%	15.5%	25.6%	14.6%	23.1%	20.2%
[75, +]	7.7%	7.0%	11.8%	3.9%	30.6%	9.8%
TOTAL	100%	100%	100%	100%	100%	100%

* Categories [75,80[and [80,+[are merged to avoid cells with frequencies below 5.

Education (N=1975)	**CM-1**	**CM-2**	**CM-3**	**CM-4**	**CM-5**	**Column total**
Primary or less	16.8%	23.4%	24.1%	13.7%	42.6%	21.6%
Secondary	35.3%	40.3%	40.2%	35.3%	29.7%	37.3%
Tertiary	48.0%	36.3%	35.7%	51.0%	27.7%	41.1%
TOTAL	100%	100%	100%	100%	100%	100%

Income (N=2010)	**CM-1**	**CM-2**	**CM-3**	**CM-4**	**CM-5**	**Column total**
Above the average	51.0%	52.8%	43.1%	58.3%	26.3%	48.8%
Similar to average	15.4%	15.5%	19.5%	13.9%	24.4%	17.0%
Below the average	10.1%	9.4%	12.5%	9.0%	18.7%	11.0%
Not declared	23.5%	22.3%	24.9%	18.8%	30.6%	23.2%
TOTAL	100%	100%	100%	100%	100%	100%

Employment (N=1985)	**CI-1**	**CI-2**	**CI-3**	**CI-4**	**CI-5**	**Column total**
Active	21.7%	23.8%	16.5%	30.8%	8.9%	21.6%
Inactive	78.3%	76.2%	83.5%	69.2%	91.1%	78.4%
TOTAL	100%	100%	100%	100%	100%	100%

Table A.13. Results of media with internet cluster (The Netherlands)

Class	N	Percentage	tvset (%)	radset (%)	newsprint (%)	booksprint (%)	newsint (%)	bookselec (%)	internews (%)	interemails (%)	intergames (%)	intersns (%)	intershopping (%)	interhobbies (%)	tvcomp (%)	interreadblogs (%)	interchat (%)	radcomp (%)
G1	366	46,04%	93,15	65,88	64,96	42,68	71,34	26,40	76,88	91,95	38,97	67,71	64,87	58,17	22,61	17,57	60,98	17,00
G2	429	53,96%	93,03	64,07	72,58	44,95	20,51	10,04	30,49	71,52	32,93	42,36	26,32	22,25	5,81	4,07	38,32	4,33
TOTAL	795	54%	93,09	64,90	69,07	43,90	43,91	17,57	51,85	80,93	35,71	54,03	44,07	38,79	13,54	10,29	48,75	10,16

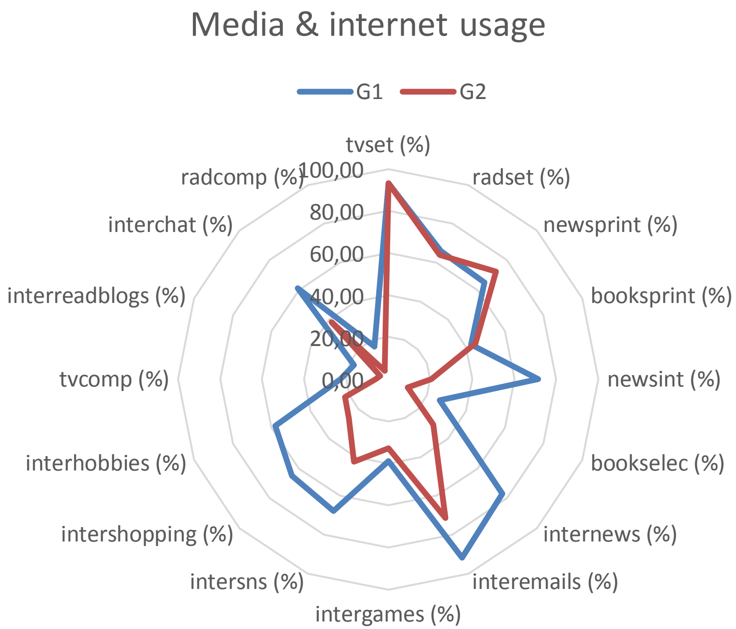


Table.A.14. The Netherlands chi-square test results (α=0.05)

THE NETHERLANDS	CI	CM
Gender	p = 0.03137023 Minimum expected frequency: 173.1019	p = 0.5081167 Minimum expected frequency: 97.82507
Age 5 categories	p-value = 0.6021006 Minimum expected frequency: 16.11321	p=9.753e-05 Minimum expected frequency: 9.458225
Has partner	p = 0.01530088 Minimum expected frequency: 50.92172	p=0.000308544 Minimum expected frequency: 56.15727
Has children	p = 0.858274 Minimum expected frequency: 45.78283	p=0.07447005 Minimum expected frequency: 50.75754
Educational level	p = 0.001653541 Minimum expected frequency: 44.61274	p=0.1676458 Minimum expected frequency: 24.64333
Income	p-value = 0.04860519 Minimum expected frequency: 58.46792	p=0.0302 Minimum expected frequency: 33.50914
Employment	p = 0.3428544 Minimum expected frequency: 11.26904	p = 3.42853e-05 Minimum expected frequency: 12.48485
CM	p=4.6029e-10 Minimum expected frequency: 48.37206	-

Table A.15. Cross-table outputs of the Netherlands media & internet usage

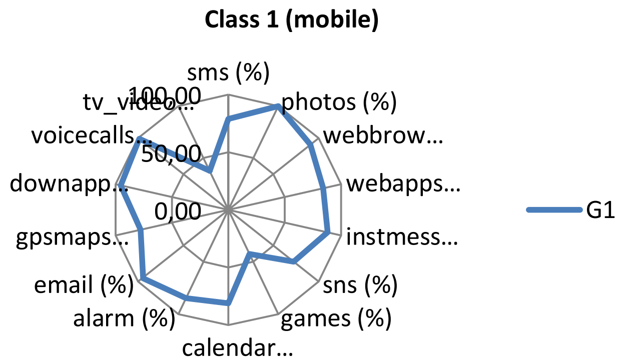
Gender (N=795)	**CI-1**	**CI-2**	**Column total**
Male	56.83%	49,18%	52.7%
Female	43.17%	50.82%	47.3%
TOTAL	100%	100%	100%

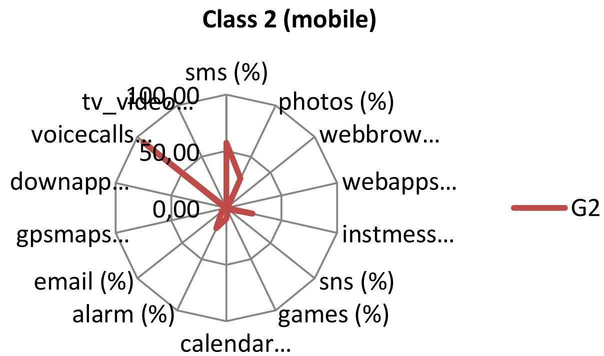
Has partner (N=792)	**CI-1**	**CI-2**	**Column total**
Married	76.65%	68.93%	72.5%
Not married	23.35%	31.07%	27.5%
TOTAL	100%	100%	100%

Education (N=785)	**CI-1**	**CI-2**	**Column total**
Primary or less	8.0%	16.1%	12.4%
Secondary	58.7%	56.4%	57.5%
Tertiary	33.3%	27.5%	30.2%
TOTAL	100%	100%	100%

Table.A.16. Results of mobile phone features cluster (The Netherlands)

Class	N	Percentage	sms (%)	photos (%)	webbrowser (%)	webapps (%)	instmess (%)	sns (%)	games (%)	calendar (%)	alarm (%)	email (%)	gpsmaps (%)	downapps (%)	voicecalls (%)	tv_video (%)
G1	207	27,02%	78,70	99,50	90,80	84,10	87,90	72,00	42,50	81,20	85,00	94,70	77,80	95,70	98,60	37,70
G2	231	30,16%	57,60	29,00	2,60	0,00	23,80	1,70	3,00	9,50	20,30	4,80	0,40	1,70	93,50	0,00
G3	328	42,82%	61,30	93,00	38,70	29,30	61,30	38,40	16,80	35,10	34,80	62,20	29,90	45,10	93,00	5,50
TOTAL	766	100%	64,89	75,45	41,89	35,27	57,18	36,41	19,58	39,84	43,99	53,67	33,95	45,68	94,66	12,54





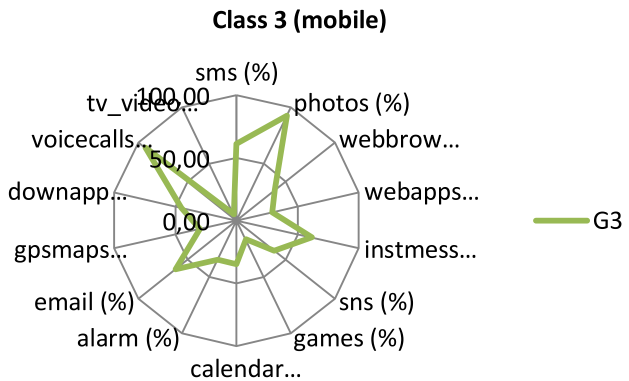


Table A.17. Cross-table outputs of the Netherlands mobile phone usage

Gender (N=766)	CM-1	CM-2	CM-3	Column total
Male	56.0%	52.4%	50.9%	52.7%
Female	44.0%	47.6%	49.1%	47.3%
TOTAL	100%	100%	100%	100%

Age in 5 categories (N=766)	CM-1	CM-2	CM-3	Column total
[60,65[	43.5%	23.8%	37.2%	34.9%
[65,70[	32.9%	32.0%	33.5%	32.9%
[70,75[	14.5%	23.8%	16.8%	18.2%
[75,80[	7.6%	13.0%	7.9%	9.4%
[80,+ [	1.5%	7.4%	4.6%	4.6%
TOTAL	100%	100%	100%	100%

Has partner (N=763)	CM-1	CM-2	CM-3	Column total
Married	83.0%	66.5%	70.6%	72.7%
Not married	17.0%	33.5%	29.4%	27.3%
TOTAL	100%	100%	100%	100%

Has children (N=763)	CM-1	CM-2	CM-3	Column total
Yes	81.1%	72.2%	74.0%	75.4%
No	18.9%	27.8%	26.0%	24.6%
TOTAL	100%	100%	100%	100%

Education (N=785)	CM-1	CM-2	CM-3	Column total
Primary or less	8.3%	15.0%	12.3%	12.0%
Secondary	58.0%	53.3%	59.7%	57.3%
Tertiary	33.7%	31.7%	28.0%	30.7%
TOTAL	100%	100%	100%	100%

Income (N=766)	CM-1	CM-2	CM-3	Column total
Above the average	36.7%	23.8%	27.4%	28.9%

Similar to average	16.4%	13.9%	17.7%	16.1%
Below the average	28.0%	34.2%	33.2%	32.1%
Not declared	18.9%	28.1%	21.7%	22.9%
TOTAL	100%	100%	100%	100%

Employ (N=759)	**CM-1**	**CM-2**	**CM-3**	**Column total**
Active	27.7%	11.0%	21.5%	20.0%
Inactive	64.0%	84.6%	71.7%	73.9%
Other	8.3%	4.4%	5.8%	6.1%
TOTAL	100%	100%	100%	100%

(N=766)	**CI-1**	**CI-2**	**CI-3**	**Column total**
CM-1	16.2%	25.7%	37.1%	27.0%
CM-2	42.3%	30.7%	19.4%	30.2%
CM-3	42.5%	43.6%	43.5%	42.8%
TOTAL	100%	100%	100%	100%

Appendix: R code

```
#################################################################
load("C:/Users/lu/Desktop/tfg/.RData")                          #
#################################################################
set.seed(1994)
#Install packages
install.packages("MASS")
install.packages(“foreign")
install.packages("survey")
install.packages("poLCA")
install.packages("memisc")
install.packages("moments")
install.packages("Hmisc")
install.packages("gmodels")
install.packages entropy)
install.packages (ggplot2)
install.packages (reshape2)
install.packages (agricolae)
install.packages (gmodels)

# Activate the library
library(MASS)
library(foreign)
library(survey)
library(poLCA)
library(memisc)
library(moments)
library(Hmisc)
library(entropy)
library(ggplot2)
library(reshape2)
library(agricolae)
```

```
library(gmodels)

#Set up the directory
setwd("C:/Users/lu/Desktop/tfg")

# Read the SPSS data
mySPSSData <- read.spss("TFG_LuLi_OlderAudiences_Feb2018_3Countries.sav",
                        to.data.frame=TRUE,
                        use.value.labels=F)

View(mySPSSData)
dim(mySPSSData)
colnames(mySPSSData)
##6577obs of 300 variables
#Eliminate the irrelevant variables of this study
bbdd<-mySPSSData[,-
c(2,4:5,8:9,12:13,16:17,20:21,24:25,28:29,32:33,36:37,40:41,44:45,48:74,95:141,142:143,146:147,150:1
51,154:155,158:159,162:163,166:167,170:171,174:175,178:179,182:183,186:232,247:300)]
attach(bbdd)
dim(bbdd)
#6577obs of 80 variables

##########################Build the new variables (dichotomous)
###########Yes - Other, where Yes=Used, Other = Not used or Didn't remember
#question 1
bbdd$q1_tvset <- NA
bbdd$q1_tvset[bbdd$Q1_TV_TVSET_NO==0 &bbdd$Q1_TV_TVSET_REM==0]<-1
bbdd$q1_tvset[bbdd$Q1_TV_TVSET_NO==1&bbdd$Q1_TV_TVSET_REM==0]<-0
bbdd$q1_tvset[bbdd$Q1_TV_TVSET_NO==0&bbdd$Q1_TV_TVSET_REM==1]<-0
bbdd$q1_tvset <- factor(bbdd$q1_tvset, levels = c(1,0),
                        labels = c("Yes", "other"  ))


bbdd$q1_tvcomp <- NA
bbdd$q1_tvcomp[bbdd$Q1_TV_COMP_NO==0 &bbdd$Q1_TV_COMP_REM==0 ]<-1
bbdd$q1_tvcomp[bbdd$Q1_TV_COMP_NO==1&bbdd$Q1_TV_COMP_REM==0]<-0
bbdd$q1_tvcomp[bbdd$Q1_TV_COMP_NO==0&bbdd$Q1_TV_COMP_REM==1]<-0
bbdd$q1_tvcomp <- factor(bbdd$q1_tvcomp, levels = c(1,0),
                    labels = c("Yes", "other"  ))
bbdd$q1_tvmob <- NA
bbdd$q1_tvmob[bbdd$Q1_TV_MOB_NO==0 &bbdd$Q1_TV_MOB_REM==0 ]<-1
bbdd$q1_tvmob[bbdd$Q1_TV_MOB_NO==1&bbdd$Q1_TV_MOB_REM==0]<-0
bbdd$q1_tvmob[bbdd$Q1_TV_MOB_NO==0&bbdd$Q1_TV_MOB_REM==1]<-0
bbdd$q1_tvmob <- factor(bbdd$q1_tvmob, levels = c(1,0),
                   labels = c("Yes", "other"  ))


bbdd$q1_radset <- NA
bbdd$q1_radset[bbdd$Q1_RAD_SET_NO==0 &bbdd$Q1_RAD_SET_REM==0 ]<-1
bbdd$q1_radset[bbdd$Q1_RAD_SET_NO==1&bbdd$Q1_RAD_SET_REM==0]<-0
bbdd$q1_radset[bbdd$Q1_RAD_SET_NO==0&bbdd$Q1_RAD_SET_REM==1]<-0
bbdd$q1_radset<- factor(bbdd$q1_radset, levels = c(1,0),
                   labels = c("Yes", "other"  ))


bbdd$q1_radmob <- NA
bbdd$q1_radmob[bbdd$Q1_RAD_MOB_NO==0 &bbdd$Q1_RAD_MOB_REM==0 ]<-1
bbdd$q1_radmob[bbdd$Q1_RAD_MOB_NO==1&bbdd$Q1_RAD_MOB_REM==0]<-0
bbdd$q1_radmob[bbdd$Q1_RAD_MOB_NO==0&bbdd$Q1_RAD_MOB_REM==1]<-0
bbdd$q1_radmob<- factor(bbdd$q1_radmob, levels = c(1,0), labels = c("Yes", "other"  ))
```

```
bbdd$q1_radcomp<-NA
bbdd$q1_radcomp[bbdd$Q1_RAD_COMP_NO==0 &bbdd$Q1_RAD_COMP_REM==0]<-1
bbdd$q1_radcomp[bbdd$Q1_RAD_COMP_NO==1&bbdd$Q1_RAD_COMP_REM==0]<-0
bbdd$q1_radcomp[bbdd$Q1_RAD_COMP_NO==0&bbdd$Q1_RAD_COMP_REM==1]<-0
bbdd$q1_radcomp <- factor(bbdd$q1_radcomp, levels = c(1,0),
                          labels = c("Yes", "other"))

bbdd$q1_newsint <- NA
bbdd$q1_newsint[bbdd$Q1_NEWSP_INT_NO==0 &bbdd$Q1_NEWSP_INT_REM==0 ]<-1
bbdd$q1_newsint[bbdd$Q1_NEWSP_INT_NO==1&bbdd$Q1_NEWSP_INT_REM==0]<-0
bbdd$q1_newsint[bbdd$Q1_NEWSP_INT_NO==0&bbdd$Q1_NEWSP_INT_REM==1]<-0
bbdd$q1_newsint<- factor(bbdd$q1_newsint, levels = c(1,0),
                        labels = c("Yes", "other"  ))

bbdd$q1_newsprint <- NA
bbdd$q1_newsprint[bbdd$Q1_NEWSP_PRINT_NO==0 &bbdd$Q1_NEWSP_PRINT_REM==0 ]<-1
bbdd$q1_newsprint[bbdd$Q1_NEWSP_PRINT_NO==1&bbdd$Q1_NEWSP_PRINT_REM==0]<-0
bbdd$q1_newsprint[bbdd$Q1_NEWSP_PRINT_NO==0&bbdd$Q1_NEWSP_PRINT_REM==1]<-0
bbdd$q1_newsprint<- factor(bbdd$q1_newsprint, levels = c(1,0),
                         labels = c("Yes", "other"  ))

bbdd$q1_booksprint <- NA
bbdd$q1_booksprint[bbdd$Q1_BOOKS_PRINT_NO==0 &bbdd$Q1_BOOKS_PRINT_REM==0 ]<-1
bbdd$q1_booksprint[bbdd$Q1_BOOKS_PRINT_NO==1&bbdd$Q1_BOOKS_PRINT_REM==0]<-0
bbdd$q1_booksprint[bbdd$Q1_BOOKS_PRINT_NO==0&bbdd$Q1_BOOKS_PRINT_REM==1]<-0
bbdd$q1_booksprint<- factor(bbdd$q1_booksprint, levels = c(1,0),
                          labels = c("Yes", "other"  ))

bbdd$q1_bookselec <- NA
bbdd$q1_bookselec[bbdd$Q1_BOOKS_ELEC_NO==0 &bbdd$Q1_BOOKS_ELEC_REM==0 ]<-1
bbdd$q1_bookselec[bbdd$Q1_BOOKS_ELEC_NO==1&bbdd$Q1_BOOKS_ELEC_REM==0]<-0
bbdd$q1_bookselec[bbdd$Q1_BOOKS_ELEC_NO==0&bbdd$Q1_BOOKS_ELEC_REM==1]<-0
bbdd$q1_bookselec<- factor(bbdd$q1_bookselec, levels = c(1,0),
                          labels = c("Yes", "other"  ))

bbdd$q1_audbooks <- NA
bbdd$q1_audbooks[bbdd$Q1_AUDBOOKS_NO==0 &bbdd$Q1_AUDBOOKS_REM==0 ]<-1
bbdd$q1_audbooks[bbdd$Q1_AUDBOOKS_NO==1&bbdd$Q1_AUDBOOKS_REM==0]<-0
bbdd$q1_audbooks[bbdd$Q1_AUDBOOKS_NO==0&bbdd$Q1_AUDBOOKS_REM==1]<-0
bbdd$q1_audbooks<- factor(bbdd$q1_audbooks, levels = c(1,0),
                         labels = c("Yes", "other"  ))

bbdd$q15_internews <- NA
bbdd$q15_internews[bbdd$Q15_Inter_news_NO==0 &bbdd$Q15_Inter_news_REM==0 ]<-1
bbdd$q15_internews[bbdd$Q15_Inter_news_NO==1&bbdd$Q15_Inter_news_REM==0]<-0
bbdd$q15_internews[bbdd$Q15_Inter_news_NO==0&bbdd$Q15_Inter_news_REM==1]<-0
bbdd$q15_internews<- factor(bbdd$q15_internews, levels = c(1,0),
                          labels = c("Yes", "other"  ))

bbdd$q15_interemails <- NA
bbdd$q15_interemails[bbdd$Q15_Inter_emails_NO==0 &bbdd$Q15_Inter_emails_REM==0 ]<-1
bbdd$q15_interemails[bbdd$Q15_Inter_emails_NO==1&bbdd$Q15_Inter_emails_REM==0]<-0
bbdd$q15_interemails[bbdd$Q15_Inter_emails_NO==0&bbdd$Q15_Inter_emails_REM==1]<-0
bbdd$q15_interemails<- factor(bbdd$q15_interemails, levels = c(1,0),
                            labels = c("Yes", "other"  ))

bbdd$q15_interpodcast <- NA
bbdd$q15_interpodcast[bbdd$Q15_Inter_podcast_NO==0 &bbdd$Q15_Inter_podcast_REM==0 ]<-1
bbdd$q15_interpodcast[bbdd$Q15_Inter_podcast_NO==1&bbdd$Q15_Inter_podcast_REM==0]<-0
```

```
bbdd$q15_interpodcast[bbdd$Q15_Inter_podcast_NO==0&bbdd$Q15_Inter_podcast_REM==1]<-0
bbdd$q15_interpodcast<- factor(bbdd$q15_interpodcast, levels = c(1,0),
                        labels = c("Yes", "other"  ))

bbdd$q15_intergames <- NA
bbdd$q15_intergames[bbdd$Q15_Inter_games_NO==0 &bbdd$Q15_Inter_games_REM==0 ]<-1
bbdd$q15_intergames[bbdd$Q15_Inter_games_NO==1&bbdd$Q15_Inter_games_REM==0]<-0
bbdd$q15_intergames[bbdd$Q15_Inter_games_NO==0&bbdd$Q15_Inter_games_REM==1]<-0
bbdd$q15_intergames<- factor(bbdd$q15_intergames, levels = c(1,0),
                       labels = c("Yes", "other"  ))

bbdd$q15_InterSNS <- NA
bbdd$q15_InterSNS[bbdd$Q15_Inter_SNS_NO==0 &bbdd$Q15_Inter_SNS_REM==0 ]<-1
bbdd$q15_InterSNS[bbdd$Q15_Inter_SNS_NO==1&bbdd$Q15_Inter_SNS_REM==0]<-0
bbdd$q15_InterSNS[bbdd$Q15_Inter_SNS_NO==0&bbdd$Q15_Inter_SNS_REM==1]<-0
bbdd$q15_InterSNS<- factor(bbdd$q15_InterSNS, levels = c(1,0),
                     labels = c("Yes", "other"  ))


bbdd$q15_interchat <- NA
bbdd$q15_interchat[bbdd$Q15_Inter_chat_NO==0 &bbdd$Q15_Inter_chat_REM==0 ]<-1
bbdd$q15_interchat[bbdd$Q15_Inter_chat_NO==1&bbdd$Q15_Inter_chat_REM==0]<-0
bbdd$q15_interchat[bbdd$Q15_Inter_chat_NO==0&bbdd$Q15_Inter_chat_REM==1]<-0
bbdd$q15_interchat<- factor(bbdd$q15_interchat, levels = c(1,0),
                      labels = c("Yes", "other"  ))

bbdd$q15_interreadblogs <- NA
bbdd$q15_interreadblogs[bbdd$Q15_Inter_readblogs_NO==0 &bbdd$Q15_Inter_readblogs_REM==0
]<-1
bbdd$q15_interreadblogs[bbdd$Q15_Inter_readblogs_NO==1&bbdd$Q15_Inter_readblogs_REM==0]<-
0
bbdd$q15_interreadblogs[bbdd$Q15_Inter_readblogs_NO==0&bbdd$Q15_Inter_readblogs_REM==1]<-
0
bbdd$q15_interreadblogs<- factor(bbdd$q15_interreadblogs, levels = c(1,0),
                          labels = c("Yes", "other"  ))


bbdd$q15_interwriteblogs<- NA
bbdd$q15_interwriteblogs[bbdd$Q15_Inter_writeblogs_NO==0 &bbdd$Q15_Inter_writeblogs_REM==0
]<-1
bbdd$q15_interwriteblogs[bbdd$Q15_Inter_writeblogs_NO==1&bbdd$Q15_Inter_writeblogs_REM==0
]<-0
bbdd$q15_interwriteblogs[bbdd$Q15_Inter_writeblogs_NO==0&bbdd$Q15_Inter_writeblogs_REM==1
]<-0
bbdd$q15_interwriteblogs<- factor(bbdd$q15_interwriteblogs, levels = c(1,0),
                           labels = c("Yes", "other"  ))

bbdd$q15_intershopping<- NA
bbdd$q15_intershopping[bbdd$Q15_Inter_shopping_NO==0 &bbdd$Q15_Inter_shopping_REM==0 ]<-
1
bbdd$q15_intershopping[bbdd$Q15_Inter_shopping_NO==1&bbdd$Q15_Inter_shopping_REM==0]<-0
bbdd$q15_intershopping[bbdd$Q15_Inter_shopping_NO==0&bbdd$Q15_Inter_shopping_REM==1]<-0
bbdd$q15_intershopping<- factor(bbdd$q15_intershopping, levels = c(1,0),
                         labels = c("Yes", "other"  ))

bbdd$q15_interhobbies <- NA
bbdd$q15_interhobbies[bbdd$Q15_Inter_hobbies_NO==0 &bbdd$Q15_Inter_hobbies_REM==0 ]<-1
bbdd$q15_interhobbies[bbdd$Q15_Inter_hobbies_NO==1&bbdd$Q15_Inter_hobbies_REM==0]<-0
bbdd$q15_interhobbies[bbdd$Q15_Inter_hobbies_NO==0&bbdd$Q15_Inter_hobbies_REM==1]<-0
bbdd$q15_interhobbies<- factor(bbdd$q15_interhobbies, levels = c(1,0),
```

```
                        labels = c("Yes", "other"  ))


bbdd$q15_interother <- NA
bbdd$q15_interother[bbdd$Q15_Inter_other_NO==0 &bbdd$Q15_Inter_other_REM==0 ]<-1
bbdd$q15_interother[bbdd$Q15_Inter_other_NO==1&bbdd$Q15_Inter_other_REM==0]<-0
bbdd$q15_interother[bbdd$Q15_Inter_other_NO==0&bbdd$Q15_Inter_other_REM==1]<-0
bbdd$q15_interother<- factor(bbdd$q15_interother, levels = c(1,0),
                       labels = c("Yes", "other"  ))


#Add labels for the mobile usage variables
bbdd$Q7_SMS<- factor(bbdd$Q7_SMS, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_MMS<- factor(bbdd$Q7_MMS, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_TV_Video<- factor(bbdd$Q7_TV_Video, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_Radio<- factor(bbdd$Q7_Radio, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_Podcast<- factor(bbdd$Q7_Podcast, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_Music_player<- factor(bbdd$Q7_Music_player, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_Photos<- factor(bbdd$Q7_Photos, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_Rec_Video<- factor(bbdd$Q7_Rec_Video, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_Web_Browser<- factor(bbdd$Q7_Web_Browser, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_Web_Apps<- factor(bbdd$Q7_Web_Apps, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_Inst_mess<- factor(bbdd$Q7_Inst_mess, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_SNS<- factor(bbdd$Q7_SNS, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_Games<- factor(bbdd$Q7_Games, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_Calendar<- factor(bbdd$Q7_Calendar, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_Alarm<- factor(bbdd$Q7_Alarm, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_Email<- factor(bbdd$Q7_Email, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_GPS_Mapps<- factor(bbdd$Q7_GPS_Mapps, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_Down_Apps<- factor(bbdd$Q7_Down_Apps, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_Voice_calls<- factor(bbdd$Q7_Voice_calls, levels = c(1,0), labels = c("Yes", "No"  ))
bbdd$Q7_Other<- factor(bbdd$Q7_Other, levels = c(1,0), labels = c("Yes", "No"  ))

#Recode the social demographic variables
bbdd$Q22_Age_5cat<-NA
bbdd$Q22_Age_5cat[bbdd$Q22_Age < 65] <- 1
bbdd$Q22_Age_5cat[bbdd$Q22_Age >= 65 & bbdd$Q22_Age < 70 ] <- 2
bbdd$Q22_Age_5cat[bbdd$Q22_Age >= 70 & bbdd$Q22_Age < 75 ] <- 3
bbdd$Q22_Age_5cat[bbdd$Q22_Age >= 75 & bbdd$Q22_Age < 80 ] <- 4
bbdd$Q22_Age_5cat[bbdd$Q22_Age >= 80] <- 5

bbdd$Q22_Age_5cat <- factor(bbdd$Q22_Age_5cat, levels = c(1,2,3,4,5),
                      labels = c("[60, 65[", "[65, 70[", "[70, 75[",
                             "[75, 80[", "[80, + ]"))

bbdd$has_partner<-NA
bbdd$has_partner[bbdd$Q23_Family_status != 3 & bbdd$Q23_Family_status != 4 ] <- 2
bbdd$has_partner[bbdd$Q23_Family_status == 3 | bbdd$Q23_Family_status == 4 ] <- 1
bbdd$has_partner[bbdd$Q23_Family_status == 99 ] <- 3

bbdd$has_partner  <- factor(bbdd$has_partner, levels = c(1, 2, 3),
                          labels = c("Married", "Not married", "NR"))

bbdd$has_children<-NA
bbdd$has_children[bbdd$Q23_Family_status==2]<-1
bbdd$has_children[bbdd$Q23_Family_status==4]<-1
bbdd$has_children[bbdd$Q23_Family_status==6]<-1
bbdd$has_children[bbdd$Q23_Family_status==8]<-1

bbdd$has_children[bbdd$Q23_Family_status==1]<-0
```

```
bbdd$has_children[bbdd$Q23_Family_status==3]<-0
bbdd$has_children[bbdd$Q23_Family_status==5]<-0
bbdd$has_children[bbdd$Q23_Family_status==7]<-0
bbdd$has_children[bbdd$Q23_Family_status==99]<-2
bbdd$has_children<- factor(bbdd$has_children, levels = c(1,0,2),
                    labels = c("Yes", "No","NR"  ))

bbdd$edu_3cat<-NA
bbdd$edu_3cat[bbdd$Q24_Education == 1 | bbdd$Q24_Education == 2]<- 1
bbdd$edu_3cat[bbdd$Q24_Education == 3 | bbdd$Q24_Education == 4 | bbdd$Q24_Education == 5]<-
2
bbdd$edu_3cat[bbdd$Q24_Education == 6 | bbdd$Q24_Education == 7 | bbdd$Q24_Education == 8]<-
3
bbdd$edu_3cat[bbdd$Q24_Education == 9 ] <- 4

bbdd$edu_3cat <- factor(bbdd$edu_3cat, levels = c(1,2,3,4),
                       labels = c("Primary or less", "Secondary", "Tertiary", "Don't know"))

bbdd$income_3cat<-NA
bbdd$income_3cat[bbdd$Q25_Income == 1 | bbdd$Q25_Income == 2] <- 1
bbdd$income_3cat[bbdd$Q25_Income == 3] <- 2
bbdd$income_3cat[bbdd$Q25_Income == 4 | bbdd$Q25_Income == 5] <- 3
bbdd$income_3cat[bbdd$Q25_Income == 6 | bbdd$Q25_Income == 7] <- 4

bbdd$income_3cat <- factor(bbdd$income_3cat, levels = c(1,2,3,4),
                      labels = c("Above the average", "Similar to average",
                             "Below the average", "Not declared"))

bbdd$income_2cat<-NA
bbdd$income_2cat[bbdd$Q25_Income == 1 | bbdd$Q25_Income == 2] <- 1
bbdd$income_2cat[bbdd$Q25_Income == 3 | bbdd$Q25_Income == 4 | bbdd$Q25_Income == 5 ] <- 2
bbdd$income_2cat[bbdd$Q25_Income == 6 | bbdd$Q25_Income == 7] <- 3

bbdd$income_2cat <- factor(bbdd$income_2cat, levels = c(1,2,3),
                      labels = c("Above the average", "Similar or below average",
                             "Not declared"))
bbdd$employ_3cat<-NA
bbdd$employ_3cat[bbdd$Q26_Employment == 1 | bbdd$Q26_Employment == 2 ]<- 1
bbdd$employ_3cat[bbdd$Q26_Employment == 3 | bbdd$Q26_Employment == 4|
bbdd$Q26_Employment == 5]<- 2
bbdd$employ_3cat[bbdd$Q26_Employment == 6]<- 3
bbdd$employ_3cat[bbdd$Q26_Employment == 7| bbdd$Q26_Employment == 99]<- 4


bbdd$employ_3cat <- factor(bbdd$employ_3cat, levels = c(1,2,3,4),labels = c("Active", "Inactive",
                             "Other", "DK_NR"))

#########################Separate the 3 countries##################
levels(CountryCode)
str(CountryCode)
bbdd$CountryCode1[bbdd$CountryCode==7]<-"Spain"
bbdd$CountryCode1[bbdd$CountryCode==5]<-"Holland"
bbdd$CountryCode1[bbdd$CountryCode==2]<-"Canada"
ldf <- split(bbdd, bbdd$CountryCode1)
#3538 obs in Canada
canada<-ldf$"Canada"
#801 obs in Holland
holland<-ldf$"Holland"
# 2238 obs in Spain
spain<-ldf$"Spain"
```

```
###################Histogram & boxplot for the age variable ###################
hist_boxplot2 <- function(x, title, unit) {
  #print(attributes(x[[1]]))
  #plot_title<-strsplit(attributes(x[[1]])$annotation[[1]], " - ", fixed=T)[[1]][1]
  #print(description(x)[[1]])
  plot_title<-title
  plot_title<-gsub('(.{1,42})(\\s|$)', '\\1\n', plot_title)
  layout(matrix(seq(2)),heights=c(0.85,0.25))
  # Note that the exact settings of various graphical parameters is determined
  # by a lot of trial and error.  Part of why it's nice to have things scripted.
  # Here are the rest of the plotting commands.
  options(repr.plot.width=5, repr.plot.height=4)
  par(mar=c(0,3,4,1))                    # reduce size of lower margin
  hist(x[[1]], breaks=50, axes=FALSE,         # plot a histogram with 50 bins
       main=plot_title,                            # use this title
       xlab="", ylab="Frequency", col="darkgray", border="white")#, border=NA) axis labels, color
  plot_text <- function(text, location="top"){
    legend(location,legend=text, bty ="n", pch=NA)
  }

  #error <- qt(0.975,df=length(x[[1]][!is.na(x[[1]])])-1)*sd(x[[1]],
na.rm=T)/sqrt(length(x[[1]][!is.na(x[[1]])]))
  x_mean=mean(x[[1]], na.rm=T)
  x_out=boxplot.stats(x[[1]])$out
  x_nout=x[[1]]
  x_nout=x_nout[!x_nout %in% x_out]
  print(sum(!is.na(x_nout)))
  print(x_out)
  x_noutmean=mean(x_nout, na.rm=T)
  x_sd=sd(x[[1]], na.rm=T)
  n_n=N=sum(!is.na(x[[1]]))

  #abline(v = x_mean, col = "blue", lty=2)
  #abline(v = round(mean(x[[1]], na.rm=T)-error), col = "blue", lty=3)
  #abline(v = round(mean(x[[1]], na.rm=T)+error), col = "blue", lty=3)
  axis(2)
  par(mar=c(3,3,0,1), mgp=c(2,0.5,0.0))      # adjust margins and axis location
  boxplot(x[[1]], horizontal=TRUE, axes=FALSE)  # add a boxplot underneath
  axis(1, main="A")                            # add the x-axis
  #abline(v = x_mean, col = "blue", lty=2)
  mtext(unit, side=1,line=1.5, font=2)  # add the x-axis label
  #par(old_par)
  #mtext(text='Test',side=1,line=1)
}

hist_boxplot2(Canada["Q22_Age"], "Canada_Age","Years")
hist_boxplot2(spain["Q22_Age"], "Spain_Age","Years")
hist_boxplot2(holland["Q22_Age"], "The Netherland_Age","Years")

#########################Outliers in Age (Canada)###############
summary(canada$Q22_Age)
cz<-abs(scale(canada$Q22_Age))
cout<-canada$Q22_Age[cz>3]

#return the index of the outlier
which(canada$Q22_Age %in% c(90, 95, 92, 93, 89,94,91))
Canada<-canada[-c(45,  204,  305,  335,  921, 1139, 1152 ,1522, 1817 ,1986 ,2558,
                2687, 2697, 2845, 3153 ,3168),]
#16 outliers
```

```
boxplot(canada$Q22_Age)
boxplot(Canada$Q22_Age)

################outliers in Age (Holland)####################
###cut off value:3
hz<-abs(scale(holland$Q22_Age))
hout<-holland$Q22_Age[hz>3]
#6 outliers: 88 93 86 88 95 87

#return the index of the outlier
which(holland$Q22_Age %in% c(86,87,88,93,95))

Holland<-holland[-c(86, 163, 300,390, 543, 772),]
boxplot(Holland$Q22_Age)

#################outliers in Age (Spain) ###################
summary(spain$Q22_Age)
sz<-abs(scale(spain$Q22_Age))
sout<-spain$Q22_Age[sz>3]
#14 outliers: 84  86  84  84  86  87  89  93  90 101  84  86  88  88
#return the index of the outlier
which(spain$Q22_Age %in% c(86,84,87, 89 , 93,  90, 101, 88))

Spain<-spain[-c(273,  427,  434,  511,  580,  646,  927, 1076, 1171, 1493, 1784,
        1843, 1853, 1958),]

boxplot(spain$Q22_Age)
boxplot(Spain$Q22_Age)

##########Sample size after eliminating the outliers
#Canada    3522 obs of 134 var
#Spain     2224 obs of 134 var
#Holland   795  obs of 134 var


##########IQR method
#Canada
quantiles <- quantile(canada$Q22_Age, probs = c(.25, .75))
range <- 1.5 * IQR(canada$Q22_Age)
normal_gdp <- subset( canada$Q22_Age ,
           canada$Q22_Age > (quantiles[1] - range) & canada$Q22_Age < (quantiles[2] + range))
summary(normal_gdp)
cana_out<-canada$Q22_Age[canada$Q22_Age>=quantiles[2] + range]

##Spain
quantiles <- quantile(spain$Q22_Age, probs = c(.25, .75))
range <- 1.5 * IQR(spain$Q22_Age)
normal_gdp <- subset( spain$Q22_Age,
           spain$Q22_Age > (quantiles[1] - range) & spain$Q22_Age < (quantiles[2] + range))

summary(normal_gdp)
spain_out<-spain$Q22_Age[spain$Q22_Age>=quantiles[2] + range]

#The Netherlands
quantiles <- quantile(holland$Q22_Age, probs = c(.25, .75))
range <- 1.5 * IQR(holland$Q22_Age)
normal_gdp <- subset( holland$Q22_Age ,
           holland$Q22_Age > (quantiles[1] - range) & holland$Q22_Age < (quantiles[2] + range))

summary(normal_gdp)
```

```
table( holland$Q22_Age > (quantiles[1] - range) & holland$Q22_Age < (quantiles[2] + range))
hol_out<-holland$Q22_Age[holland$Q22_Age>=quantiles[2] + range]

########################Weighting cases of Canada############

#we set ids = ~ 1 to indicate that all respondents originated from the same cluster.

canada.w <- svydesign(ids = ~1, data = Canada, weights = Canada$POND)
summary(canada.w)


#Here is a comparison of the sex ratios in the unweighted and the weighted data frames
prop.table(table(Canada$Q21_Sex))
prop.table(svytable(~Q21_Sex, design = canada.w))

prop.table(table(Canada$Q22_Age))
prop.table(svytable(~Q22_Age, design = canada.w))


##################Tables of proportion of each variable#########
####Canada weighted data

#Q1
round(prop.table(svytable(~ q1_tvcomp    , design = canada.w)),4)
round(prop.table(svytable(~q1_tvset  , design = canada.w)),4)
round(prop.table(svytable(~q1_tvmob  , design = canada.w)),4)
round(prop.table(svytable(~q1_radset  , design = canada.w)),4)
round(prop.table(svytable(~q1_radcomp   , design = canada.w)),4)
round(prop.table(svytable(~q1_radmob    , design = canada.w)),4)
round(prop.table(svytable(~q1_newsprint , design = canada.w)),4)
round(prop.table(svytable(~q1_newsint   , design = canada.w)),4)
round(prop.table(svytable(~q1_booksprint   , design = canada.w)),4)
round(prop.table(svytable(~q1_bookselec  , design = canada.w)),4)
round(prop.table(svytable(~q1_audbooks  , design = canada.w)),4)

#Internet usage
round(prop.table(svytable(~q15_internews     , design = canada.w)),4)
round(prop.table(svytable(~q15_interemails     , design = canada.w)),4)
round(prop.table(svytable(~q15_interpodcast   , design = canada.w)),4)
round(prop.table(svytable(~q15_intergames     , design = canada.w)),4)
round(prop.table(svytable(~q15_InterSNS       , design = canada.w)),4)
round(prop.table(svytable(~q15_interchat      , design = canada.w)),4)
round(prop.table(svytable(~q15_interreadblogs  , design = canada.w)),4)
round(prop.table(svytable(~q15_interwriteblogs, design = canada.w)),4)
round(prop.table(svytable(~q15_intershopping   , design = canada.w)),4)
round(prop.table(svytable(~q15_interhobbies , design = canada.w)),4)
round(prop.table(svytable(~ q15_interother, design = canada.w)),4)
table(Canada$Q15_Inter_other_NO)
#Q7
round(prop.table(svytable(~ Q7_SMS       , design = canada.w)),4)
round(prop.table(svytable(~ Q7_MMS     , design = canada.w)),4)
round(prop.table(svytable(~ Q7_TV_Video  , design = canada.w)),4)
round(prop.table(svytable(~ Q7_Radio  , design = canada.w)),4)
round(prop.table(svytable(~ Q7_Podcast    , design = canada.w)),4)
round(prop.table(svytable(~ Q7_Music_player , design = canada.w)),4)
round(prop.table(svytable(~ Q7_Photos  , design = canada.w)),4)
round(prop.table(svytable(~ Q7_Rec_Video  , design = canada.w)),4)
round(prop.table(svytable(~ Q7_Web_Browser  , design = canada.w)),4)
round(prop.table(svytable(~ Q7_Web_Apps   , design = canada.w)),4)
round(prop.table(svytable(~ Q7_Inst_mess    , design = canada.w)),4)
```

```
round(prop.table(svytable(~ Q7_SNS      , design = canada.w)),4)
round(prop.table(svytable(~ Q7_Games    , design = canada.w)),4)
round(prop.table(svytable(~ Q7_Calendar   , design = canada.w)),4)
round(prop.table(svytable(~ Q7_Alarm , design = canada.w)),4)
round(prop.table(svytable(~ Q7_Email    , design = canada.w)),4)
round(prop.table(svytable(~Q7_GPS_Mapps  , design = canada.w)),4)
round(prop.table(svytable(~Q7_Down_Apps   , design = canada.w)),4)
round(prop.table(svytable(~ Q7_Voice_calls   , design = canada.w)),4)
round(prop.table(svytable(~ Q7_Other   , design = canada.w)),4)

#variables sociodemographic

round(prop.table(svytable(~ Q21_Sex  , design = canada.w)),2)
round(prop.table(svytable(~ Q23_Family_status    , design = canada.w)),4)
round(prop.table(svytable(~Q24_Education  , design = canada.w)),4)
round(prop.table(svytable(~ Q25_Income, design = canada.w)),4)
round(prop.table(svytable(~ Q26_Employment   , design = canada.w)),4)
round(prop.table(svytable(~has_partner, design = canada.w)),4)
round(prop.table(svytable(~has_children, design = canada.w)),4)

####Spain
prop_table <- function(z){
  for(i in 81:109)
  print(round(prop.table(table(z[,i]))* 100, 2))
}
print(prop_table(Spain))

####The Netherlands
print(prop_table(Holland))

###############################################################################
##########################################Latent class analysis
######### CANADA
#####Media & internet usage

f <- cbind(q1_tvset,q1_radset,q1_newsprint,q1_booksprint,q1_newsint,
          q1_bookselec,q15_internews,
          q15_interemails,q15_intergames,q15_InterSNS,
          q15_intershopping,q15_interhobbies
          )~1
lca3 <- poLCA(f,Canada,nclass=3, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca4 <- poLCA(f,Canada,nclass=4, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca5 <- poLCA(f,Canada,nclass=5, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca6 <- poLCA(f,Canada,nclass=6, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca7 <- poLCA(f,Canada,nclass=7, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)

##### Mobile phone usage

g<-cbind(Q7_SMS,Q7_MMS,Q7_Music_player,Q7_Photos,Q7_Rec_Video,
        Q7_Web_Browser,Q7_Web_Apps,Q7_Inst_mess,Q7_SNS,Q7_Games,
        Q7_Calendar,Q7_Alarm,Q7_Email, Q7_GPS_Mapps,
        Q7_Down_Apps,Q7_Voice_calls)~1

lca3 <- poLCA(g,Canada,nclass=3, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
```

```
lca4 <- poLCA(g,Canada,nclass=4, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca5 <- poLCA(g,Canada,nclass=5, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca6 <- poLCA(g,Canada,nclass=6, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca7 <- poLCA(g,Canada,nclass=7, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)


###Strategy 1
complete <- function(data, desiredCols) {
  completeVec <- complete.cases(data[, desiredCols])
  return(data[completeVec, ])
}

###Strategy 2
h<- cbind(q1_tvset,q1_radset,q1_newsprint,q1_booksprint,q1_newsint,
         q1_radcomp,q1_bookselec,q15_internews,
         q15_interemails,q15_intergames,q15_InterSNS,
         q15_interreadblogs,q15_intershopping,q15_interhobbies,
         Q7_SMS,Q7_MMS,Q7_Music_player,Q7_Photos,Q7_Rec_Video,
         Q7_Web_Browser,Q7_Web_Apps,Q7_Inst_mess,Q7_SNS,Q7_Games,
         Q7_Calendar,Q7_Alarm,Q7_Email, Q7_GPS_Mapps,
         Q7_Down_Apps,Q7_Voice_calls)~1


lca3 <- poLCA(h,Canada1,nclass=3, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca4 <- poLCA(h,Canada1,nclass=4, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca5 <- poLCA(h,Canada1,nclass=5, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca6 <- poLCA(h,Canada1,nclass=6, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca7 <- poLCA(h,Canada1,nclass=7, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)

###########################LCA for Spain
#####Media & internet usage

s <- cbind(q1_tvset,q1_radset,q1_newsprint,q1_booksprint,q1_newsint,
          q1_radcomp,q1_bookselec,q1_tvcomp,q15_internews,
          q15_interemails,q15_intergames,q15_InterSNS,
          q15_interreadblogs,q15_intershopping,q15_interhobbies,q15_interchat
)~1
lca3 <- poLCA(s,Spain,nclass=3, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca4 <- poLCA(s,Spain,nclass=4, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca5 <- poLCA(s,Spain,nclass=5, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca6 <- poLCA(s,Spain,nclass=6, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca7 <- poLCA(s,Spain,nclass=7, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)

##### Mobile phone usage

p<-cbind(Q7_SMS,Q7_MMS,Q7_Music_player,Q7_Photos,Q7_Rec_Video,
```

```
        Q7_Web_Browser,Q7_Web_Apps,Q7_Inst_mess,Q7_SNS,Q7_Games,
        Q7_Calendar,Q7_Alarm,Q7_Email, Q7_GPS_Mapps,
        Q7_Down_Apps,Q7_Voice_calls,Q7_Radio,Q7_TV_Video)~1

lca3 <- poLCA(p,Spain,nclass=3, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca4 <- poLCA(p,Spain,nclass=4, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca5 <- poLCA(p,Spain,nclass=5, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca6 <- poLCA(p,Spain,nclass=6, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca7 <- poLCA(p,Spain,nclass=7, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)

###########################LCA for Netherland
#####Media & internet usage
n <- cbind(q1_tvset,q1_radset,q1_newsprint,q1_booksprint,q1_newsint,
         q1_radcomp,q1_bookselec,q1_tvcomp,q15_internews,
         q15_interemails,q15_intergames,q15_InterSNS,
         q15_interreadblogs,q15_intershopping,q15_interhobbies,q15_interchat
)~1
lca2 <- poLCA(n,Holland,nclass=2, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca3 <- poLCA(n,Holland,nclass=3, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca4 <- poLCA(n,Holland,nclass=4, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca5 <- poLCA(n,Holland,nclass=5, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca6 <- poLCA(n,Holland,nclass=6, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca7 <- poLCA(n,Holland,nclass=7, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)

##### Mobile phone usage

h<-cbind(Q7_SMS,Q7_Photos,Q7_Web_Browser, Q7_Web_Apps,Q7_Inst_mess,Q7_SNS,Q7_Games,
        Q7_Calendar,Q7_Alarm,Q7_Email,
Q7_GPS_Mapps,Q7_Down_Apps,Q7_Voice_calls,Q7_TV_Video)~1

lca2 <- poLCA(h,Holland,nclass=2, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca3 <- poLCA(h,Holland,nclass=3, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca4 <- poLCA(h,Holland,nclass=4, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca5 <- poLCA(h,Holland,nclass=5, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca6 <- poLCA(h,Holland,nclass=6, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)
lca7 <- poLCA(h,Holland,nclass=7, maxiter = 40000, nrep=10, graphs=TRUE, verbose=TRUE,
calc.se=TRUE)

######Results table################################################
results <- data.frame(Modell=c("Model 3"),
                log_likelihood=lca3$llik,
                df = lca3$resid.df,
                BIC=lca3$bic,
                ABIC=  (-2*lca3$llik) + ((log((lca3$N + 2)/24)) * lca3$npar),
```

```
        CAIC = (-2*lca3$llik) + lca3$npar * (1 + log(lca3$N)),
        likelihood_ratio=lca3$Gsq)

results$Modell<-as.integer(results$Modell)
results[1,1]<-c("Model 3")
results[2,1]<-c("Model 4")
results[3,1]<-c("Model 5")
results[4,1]<-c("Model 6")
results[5,1]<-c("Model 7")
results[6,1]<-c("Model 2")

results[2,2]<-lca4$llik
results[3,2]<-lca5$llik
results[4,2]<-lca6$llik
results[5,2]<-lca7$llik
results[6,2]<-lca2$llik

results[2,3]<-lca4$resid.df
results[3,3]<-lca5$resid.df
results[4,3]<-lca6$resid.df
results[5,3]<-lca7$resid.df
results[6,3]<-lca2$resid.df

results[2,4]<-lca4$bic
results[3,4]<-lca5$bic
results[4,4]<-lca6$bic
results[5,4]<-lca7$bic
results[6,4]<-lca2$bic

results[2,5]<-(-2*lca4$llik) + ((log((lca4$N + 2)/24)) * lca4$npar) #abic
results[3,5]<-(-2*lca5$llik) + ((log((lca5$N + 2)/24)) * lca5$npar) #abic
results[4,5]<-(-2*lca6$llik) + ((log((lca6$N + 2)/24)) * lca6$npar) #abic
results[5,5]<-(-2*lca7$llik) + ((log((lca7$N + 2)/24)) * lca7$npar) #abic
results[6,5]<-(-2*lca2$llik) + ((log((lca2$N + 2)/24)) * lca2$npar) #abic

results[2,6]<- (-2*lca4$llik) + lca4$npar * (1 + log(lca4$N)) #caic
results[3,6]<- (-2*lca5$llik) + lca5$npar * (1 + log(lca5$N)) #caic
results[4,6]<- (-2*lca6$llik) + lca6$npar * (1 + log(lca6$N)) #caic
results[5,6]<- (-2*lca7$llik) + lca7$npar * (1 + log(lca7$N)) #caic
results[6,6]<- (-2*lca2$llik) + lca2$npar * (1 + log(lca2$N)) #caic

results[2,7]<-lca4$Gsq
results[3,7]<-lca5$Gsq
results[4,7]<-lca6$Gsq
results[5,7]<-lca7$Gsq
results[6,7]<-lca2$Gsq

error_prior<-entropy(lca2$P)
error_post<-mean(apply(lca2$posterior,1, entropy),na.rm = TRUE)
results[6,8]<-round(((error_prior-error_post) / error_prior),3)

error_prior<-entropy(lca3$P)
error_post<-mean(apply(lca2$posterior,1, entropy),na.rm = TRUE)
results[6,8]<-round(((error_prior-error_post) / error_prior),3)

error_prior<-entropy(lca4$P)
error_post<-mean(apply(lca4$posterior,1, entropy),na.rm = TRUE)
results[2,8]<-round(((error_prior-error_post) / error_prior),3)

error_prior<-entropy(lca5$P)
```

```
error_post<-mean(apply(lca5$posterior,1, entropy),na.rm = TRUE)
results[3,8]<-round(((error_prior-error_post) / error_prior),3)

error_prior<-entropy(lca6$P)
error_post<-mean(apply(lca6$posterior,1, entropy),na.rm = TRUE)
results[4,8]<-round(((error_prior-error_post) / error_prior),3)

results

###################Final number of classes for each cluster
##Canada
canada_media_lca5<- poLCA(f,Canada,nclass=5, maxiter = 40000, nrep=10, graphs=TRUE,
verbose=TRUE, calc.se=TRUE)
canada_mobile_lca4<-poLCA(g,Canada,nclass=4, maxiter = 40000, nrep=10, graphs=TRUE,
verbose=TRUE, calc.se=TRUE)
##Spain
Spain_media_lca5 <- poLCA(s,Spain,nclass=5, maxiter = 40000, nrep=10, graphs=TRUE,
verbose=TRUE, calc.se=TRUE)
Spain_mobile_lca5 <- poLCA(p,Spain,nclass=5, maxiter = 40000, nrep=10, graphs=TRUE,
verbose=TRUE, calc.se=TRUE)
## The Netherlands
holland_media_lca2<- poLCA(n,Holland,nclass=2, maxiter = 40000, nrep=10, graphs=TRUE,
verbose=TRUE, calc.se=TRUE)
holland_mobile_lca3 <- poLCA(h,Holland,nclass=3, maxiter = 40000, nrep=10, graphs=TRUE,
verbose=TRUE, calc.se=TRUE)

#################################Profiling
#Canada
#number of observations in each class
table(canada_media_lca5$predclass)
table(canada_mobile_lca4$predclass)

#proportion table
print(round(100*prop.table(table(canada_media_lca5$predclass)), 2))
print(round(100*prop.table(table(canada_mobile_lca4$predclass)), 2))

#Spain
#number of observations in each class
table(Spain_media_lca5$predclass)
table(Spain_mobile_lca5$predclass)

###proportion table
print(round(100*prop.table(table(Spain_media_lca5$predclass)), 2))
print(round(100*prop.table(table(Spain_mobile_lca5$predclass)), 2))

#The Netherlands
#number of observations in each class
table(holland_media_lca2$predclass)
table(holland_mobile_lca3$predclass)

###proportion table
print(round(100*prop.table(table(holland_media_lca2$predclass)), 2))
print(round(100*prop.table(table(holland_mobile_lca3$predclass)), 2))


######Univariate descriptions of socio-demographic variables for each cluster####
################Canada media usage################################
varca<-c("q1_tvset", "q1_radset", "q1_newsprint", "q1_booksprint", "q1_newsint",
        "q1_bookselec", "q15_internews", "q15_interemails",
       "q15_intergames", "q15_InterSNS", "q15_intershopping",
```

```
    "q15_interhobbies")

for (n in varca){
  print(n)
  print(round(100*prop.table(table(
    canada_media_lca5$predclass, Canada[[n]]),1), 2))
}
for (n in varca){
  print(n)
  print(CrossTable(canada_media_lca5$predclass, Canada[[n]]))
}

#GG plot
lcModel = canada_media_lca5
lcModelProbs <- melt(lcModel$probs)
str(factor(canada_media_lca5$predclass))

zp2 <- ggplot(lcModelProbs,
              aes(x = Var1, y = value, fill = Var2))
zp2 <- zp2 + geom_bar(stat = "identity", position = "stack")
zp2 <- zp2 + facet_wrap(~ L1)
zp2 <- zp2 + scale_x_discrete("Class", expand = c(0, 0))
zp2 <- zp2 + scale_y_continuous("Proportion", expand = c(0, 0))
zp2 <- zp2 + scale_fill_discrete("Factor Level")
zp2 <- zp2 + theme_bw()
print(zp2)

Canada$class = NULL
Canada$class_media = canada_media_lca5$predclass
Canada$class_media = factor(Canada$class)

#Age
boxplot(Q22_Age~class_media,data=Canada, main="Age by Class",
    xlab="Class", ylab="Age")
pairwise.t.test(Canada$Q22_Age, Canada$class_media, p.adj = "bonferroni")

#######Weighting cases in Canada
#we set ids = ~ 1 to indicate that all respondents originated from the same cluster.

canada.w <- svydesign(ids = ~1, data = Canada, weights = Canada$POND)
summary(canada.w)

#Gender
CrossTable(svytable(~class_media+Q21_Sex, design = canada.w),prop.chisq =
F,format="SPSS",chisq=T)

#Age with 5 categories
CrossTable(svytable(~class_media+Q22_Age_5cat, design = canada.w),prop.chisq =
F,format="SPSS",chisq=T)

#Income
CrossTable(svytable(~class_media+income_3cat, design = canada.w),prop.chisq =
F,format="SPSS",chisq=T)

#Family status
#Nobody prefer not to answer
#So I can exlude the "NR" level of the variable has_partner
da<-droplevels(Canada$has_partner,exclude="NR")
CrossTable(svytable(~class_media+da, design = canada.w),prop.chisq = F,format="SPSS",chisq=T)
```

```
#Has Children?
#Nobody prefer not to answer
dn<-droplevels(Canada$has_children,exclude="NR")
CrossTable(svytable(~class_media+dn, design = canada.ch),prop.chisq = F,format="SPSS",chisq=T)
#employment
cy<-Canada[!(Canada$employ_3cat=="DK_NR" | Canada$employ_3cat=="Other"),]
cy<-droplevels(cy,cy$employ_3cat)
canada.cy<- svydesign(ids = ~1, data = cy, weights = cy$POND)
CrossTable(svytable(~class_media+employ_3cat, design =canada.cy),prop.chisq =
F,format="SPSS",chisq=T)

#Education
#I decided to remove the Don't knows of the variable edu_3cat
ce<-Canada[!(Canada$edu_3cat=="Don't know" ),]
ce<-droplevels(ce,ce$edu_3cat)
canada.ce<- svydesign(ids = ~1, data = ce, weights = ce$POND)
CrossTable(svytable(~class_media+edu_3cat, design =canada.ce),prop.chisq =
F,format="SPSS",chisq=T)

################Canada mobile phone usage###################
varph<-c("Q7_SMS","Q7_MMS", "Q7_Photos", "Q7_Web_Browser", "Q7_Web_Apps",
"Q7_Inst_mess",
         "Q7_SNS", "Q7_Games", "Q7_Calendar","Q7_Alarm", "Q7_GPS_Mapps",
         "Q7_Down_Apps", "Q7_Voice_calls","Q7_Music_player",
       "Q7_Rec_Video")

for (n in varph){
  print(n)
  print(round(100*prop.table(table(
    canada_mobile_lca4$predclass, na.omit(Canada[[n]])),1), 1))
}
print(round(100*prop.table(table(
  canada_mobile_lca4$predclass, na.omit(Canada[["Q7_Email"]])),1), 1))

for (n in varph){
  print(n)
  print(CrossTable(
    canada_mobile_lca4$predclass, na.omit(Canada[[n]])))
}
##GG plot
lcModel = canada_mobile_lca4
lcModelProbs <- melt(lcModel$probs)
str(factor(canada_mobile_lca4$predclass))

zp2 <- ggplot(lcModelProbs,
              aes(x = Var1, y = value, fill = Var2))
zp2 <- zp2 + geom_bar(stat = "identity", position = "stack")
zp2 <- zp2 + facet_wrap(~ L1)
zp2 <- zp2 + scale_x_discrete("Class", expand = c(0, 0))
zp2 <- zp2 + scale_y_continuous("Proportion", expand = c(0, 0))
zp2 <- zp2 + scale_fill_discrete("Factor Level")
zp2 <- zp2 + theme_bw()
print(zp2)

complete <- function(data, desiredCols) {
  completeVec <- complete.cases(data[, desiredCols])
  return(data[completeVec, ])
}
```

```
Canada_mob<-complete(Canada,"Q7_SMS")
Canada_mob$class = NULL
Canada_mob$class =canada_mobile_lca4$predclass
Canada_mob$class = factor(Canada_mob$class)

canada.mo <- svydesign(ids = ~1, data = Canada_mob, weights = Canada_mob$POND)
summary(canada.mo)

#Age
boxplot(Q22_Age~class,data=Canada_mob, main="Age by Class",
        xlab="Class", ylab="Age")
pairwise.t.test(Canada_mob$Q22_Age, Canada_mob$class, p.adj = "bonferroni")
#test de anova
ajustee <- lm(Canada_mob$Q22_Age ~ Canada_mob$class)
anova(ajustee)

#gender
CrossTable(svytable(~class+Q21_Sex, design = canada.mo),prop.chisq = F,format="SPSS",chisq=T)

#Age 5 categories
CrossTable(svytable(~class+Q22_Age_5cat, design = canada.mo),prop.chisq =
F,format="SPSS",chisq=T)

#income
CrossTable(svytable(~class+income_3cat, design = canada.mo),prop.chisq = F,format="SPSS",chisq=T)

#Family status
#Nobody prefer not to answer
dm<-droplevels(Canada_mob$has_partner,exclude="NR")
CrossTable(svytable(~class+dm, design = canada.mo),prop.chisq = F,format="SPSS",chisq=T)

#has children?
#Nobody prefer not to answer
dq<-droplevels(Canada_mob$has_children,exclude="NR")
CrossTable(svytable(~class+dq, design = canada.mo),prop.chisq = F,format="SPSS",chisq=T)

#employment
mp<-Canada_mob[!(Canada_mob$employ_3cat=="DK_NR"| Canada_mob$employ_3cat=="Other"),]
mp<-droplevels(mp,mp$employ_3cat)
canada.mp<- svydesign(ids = ~1, data = mp, weights = mp$POND)
CrossTable(svytable(~class+employ_3cat, design = canada.mp),prop.chisq = F,format="SPSS",chisq=T)

#education
#I decided to remove the Don't knows of the variable edu_3cat
cd<-Canada_mob[!(Canada_mob$edu_3cat=="Don't know"),]
cd<-droplevels(cd,cd$edu_3cat)
canada.cd<- svydesign(ids = ~1, data = cd, weights = cd$POND)
CrossTable(svytable(~class+edu_3cat, design = canada.cd),prop.chisq = F,format="SPSS",chisq=T)


################Spain media usage##############################
varsa<-c("q1_tvset", "q1_radset", "q1_newsprint", "q1_booksprint", "q1_newsint",
        "q1_radcomp", "q1_bookselec", "q1_tvcomp", "q15_internews", "q15_interemails",
        "q15_intergames", "q15_InterSNS", "q15_interreadblogs", "q15_intershopping",
        "q15_interhobbies","q15_interchat")

for (n in varsa){
  print(n)
  print(round(100*prop.table(table(
    Spain_media_lca5$predclass, Spain[[n]]),1), 2))
```

```
}
for (n in varsa){
  print(n)
  print(CrossTable(
    Spain_media_lca5$predclass, Spain[[n]]))
}

#GG plot
lcModel = Spain_media_lca5
lcModelProbs <- melt(lcModel$probs)
str(factor(canada_media_lca5$predclass))

zp2 <- ggplot(lcModelProbs,
              aes(x = Var1, y = value, fill = Var2))
zp2 <- zp2 + geom_bar(stat = "identity", position = "stack")
zp2 <- zp2 + facet_wrap(~ L1)
zp2 <- zp2 + scale_x_discrete("Class", expand = c(0, 0))
zp2 <- zp2 + scale_y_continuous("Proportion", expand = c(0, 0))
zp2 <- zp2 + scale_fill_discrete("Factor Level")
zp2 <- zp2 + theme_bw()
print(zp2)

Spain$class = NULL
Spain$class_media = Spain_media_lca5$predclass
Spain$class_media = factor(Spain$class)

#Age
boxplot(Q22_Age~class_media,data=Spain, main="Age by Class",
        xlab="Class", ylab="Age")
pairwise.t.test(Spain$Q22_Age, Spain$class, p.adj = "bonferroni")
#test de anova
ajuste <- lm(Spain$Q22_Age ~ Spain$class_media)
anova(ajuste)


#gender
round(prop.table(table(Spain$class_media, Spain$Q21_Sex),1)*100, 1)
pairwise.t.test(Spain$Q21_Sex, Spain$class_media, p.adj = "bonferroni")
CrossTable(Spain$class_media, Spain$Q21_Sex,prop.chisq = F,format = "SPSS",digits = 2,chisq = T)

#Age with 5 categories
round(prop.table(table(Spain$class_media, Spain$Q22_Age_5cat),1)*100, 1)
CrossTable(Spain$class_media, Spain$Q22_Age_5cat,prop.chisq = F,format = "SPSS",digits = 2,chisq = 
T)
#cross table without the category [80,+]
spain_age<-Spain[!(Spain$Q22_Age_5cat=="[80, + ]"),]
spain_age<-droplevels(spain_age,spain_age$Q22_Age_5cat)
CrossTable(spain_age$class_media, spain_age$Q22_Age_5cat,prop.chisq = F,format="SPSS",chisq = T)

#Income
round(prop.table(table(Spain$class_media, Spain$income_3cat),1)*100, 1)
chisq.test(table(Spain$class_media, Spain$income_3cat))
CrossTable(Spain$income_3cat,Spain$class_media,prop.chisq = F,format = "SPSS",digits = 1,chisq = T)

#Family status
round(prop.table(table(Spain$class_media, Spain$has_partner),1)*100, 1)
spa<-Spain[!(Spain$has_partner=="NR"),]
CrossTable(spa$class_media, spa$has_partner,prop.chisq = F,format = "SPSS",digits = 2,chisq = T)

#Has Children?
```

```
round(prop.table(table(Spain$class_media, Spain$has_children),1)*100, 1)
sc<-Spain[!(Spain$has_children=="NR"),]
CrossTable(sc$class_media, sc$has_children,prop.chisq = F,format = "SPSS",digits = 2,chisq = T)

#employment
sp<-Spain[!(Spain$employ_3cat=="DK_NR" |Spain$employ_3cat=="Other"),]
CrossTable(sp$class_media, sp$employ_3cat,prop.chisq = F,format = "SPSS",digits = 1,chisq = T)

#Education
round(prop.table(table(Spain$class_media, Spain$edu_3cat),1)*100, 1)
#I decided to remove the Don't knows of the variable edu_3cat
sd<-Spain[!(Spain$edu_3cat=="Don't know"),]
round(prop.table(table(sd$class_media, sd$edu_3cat),1)*100, 1)
CrossTable(sd$class_media, sd$edu_3cat,prop.chisq = F,format = "SPSS",digits = 2,chisq = T)


################Spain mobile usage##############################

varms<-c("Q7_SMS","Q7_MMS", "Q7_Photos", "Q7_Web_Browser", "Q7_Web_Apps",
"Q7_Inst_mess",
        "Q7_SNS", "Q7_Games", "Q7_Calendar","Q7_Alarm", "Q7_GPS_Mapps", "Q7_Rec_Video",
        "Q7_Down_Apps", "Q7_Voice_calls", "Q7_TV_Video","Q7_Music_player","Q7_Radio")

for (n in varms){
  print(n)
  print(round(100*prop.table(table(
    Spain_mobile_lca5$predclass, na.omit(Spain[[n]])),1), 1))
}
print(round(100*prop.table(table(
  Spain_mobile_lca5$predclass, na.omit(Spain[["Q7_Email"]])),1), 1))

for (n in varms){
  print(n)
  print(CrossTable(
    Spain_mobile_lca5$predclass, na.omit(Spain[[n]])))
}
##GG plot
lcModel = Spain_mobile_lca5
lcModelProbs <- melt(lcModel$probs)
str(factor(Spain_mobile_lca5$predclass))

zp2 <- ggplot(lcModelProbs,
             aes(x = Var1, y = value, fill = Var2))
zp2 <- zp2 + geom_bar(stat = "identity", position = "stack")
zp2 <- zp2 + facet_wrap(~ L1)
zp2 <- zp2 + scale_x_discrete("Class", expand = c(0, 0))
zp2 <- zp2 + scale_y_continuous("Proportion", expand = c(0, 0))
zp2 <- zp2 + scale_fill_discrete("Factor Level")
zp2 <- zp2 + theme_bw()
print(zp2)

complete <- function(data, desiredCols) {
  completeVec <- complete.cases(data[, desiredCols])
  return(data[completeVec, ])
}

Spain_mob<-complete(Spain,"Q7_SMS")
Spain_mob$class = NULL
Spain_mob$class =Spain_mobile_lca5$predclass
Spain_mob$class = factor(Spain_mob$class)
```

```
#Age
boxplot(Q22_Age~class,data=Spain_mob, main="Age by Class",
       xlab="Class", ylab="Age")

pairwise.t.test(Spain_mob$Q22_Age, Spain_mob$class, p.adj = "bonferroni")
#test de anova
ajuste1 <- lm(Spain_mob$Q22_Age ~ Spain_mob$class)
anova(ajuste1)

#gender
round(prop.table(table(Spain_mob$class, Spain_mob$Q21_Sex),1)*100, 1)
pairwise.t.test(Spain_mob$Q21_Sex, Spain_mob$class, p.adj = "bonferroni")
CrossTable(Spain_mob$class,Spain_mob$Q21_Sex, prop.chisq = F,format = "SPSS",digits = 2,chisq =
T)

#Age 5 categories
chisq.test(table(Spain_mob$class, Spain_mob$Q22_Age_5cat))
CrossTable(Spain_mob$class,Spain_mob$Q22_Age_5cat, prop.chisq = F,format = "SPSS",digits = 2)
#cross table without the category [80,+]
spain_age1<-Spain_mob[!(Spain_mob$Q22_Age_5cat=="[80, + ]"),]
spain_age1<-droplevels(spain_age1,spain_age1$Q22_Age_5cat)
CrossTable(spain_age1$class, spain_age1$Q22_Age_5cat,prop.chisq = F,format="SPSS",digits=1,chisq
= T)

#income
round(prop.table(table(Spain_mob$class, Spain_mob$income_3cat),1)*100, 1)
chisq.test(table(Spain_mob$class, Spain_mob$income_3cat))
CrossTable(Spain_mob$income_3cat,Spain_mob$class, prop.chisq = F,format = "SPSS",digits = 2,chisq
= T)

#Family status
round(prop.table(table(Spain_mob$class, Spain_mob$has_partner),1)*100, 1)
st<-Spain_mob[!(Spain_mob$has_partner=="NR"),]
CrossTable(st$class, st$has_partner,prop.chisq = F,format = "SPSS",digits = 2,chisq = T)

#has children?
round(prop.table(table(Spain_mob$class, Spain_mob$has_children),1)*100, 1)
sc<-Spain_mob[!(Spain_mob$has_children=="NR"),]
CrossTable(sc$class, sc$has_children,prop.chisq = F,format = "SPSS",digits = 2,chisq = T)

#employment
sl<-Spain_mob[!(Spain_mob$employ_3cat=="DK_NR" |Spain_mob$employ_3cat=="Other"),]
CrossTable(sl$class, sl$employ_3cat,prop.chisq = F,format = "SPSS",digits = 1,chisq = T)

#education
round(prop.table(table(Spain_mob$class, Spain_mob$edu_3cat),1)*100, 1)
#I decided to remove the Don't knows of the variable edu_3cat
se<-Spain_mob[!(Spain_mob$edu_3cat=="Don't know"),]
CrossTable(se$class, se$edu_3cat,prop.chisq = F,format = "SPSS",digits = 2,chisq = T)

#################The Netherland media usage###########

varh<-c("q1_tvset", "q1_radset", "q1_newsprint", "q1_booksprint", "q1_newsint",
       "q1_radcomp", "q1_bookselec", "q1_tvcomp", "q15_internews", "q15_interemails",
       "q15_intergames", "q15_InterSNS", "q15_interreadblogs", "q15_intershopping",
       "q15_interhobbies","q15_interchat")

for (n in varh){
  print(n)
```

```
  print(round(100*prop.table(table(
    holland_media_lca3$predclass, Holland[[n]]),1), 2))
}
for (n in varh){
  print(n)
  print(CrossTable(
    holland_media_lca3$predclass, Holland[[n]]))
}


#GG plot
lcModel = holland_media_lca2
lcModelProbs <- melt(lcModel$probs)
str(factor(holland_media_lca3$predclass))

zp2 <- ggplot(lcModelProbs,
              aes(x = Var1, y = value, fill = Var2))
zp2 <- zp2 + geom_bar(stat = "identity", position = "stack")
zp2 <- zp2 + facet_wrap(~ L1)
zp2 <- zp2 + scale_x_discrete("Class", expand = c(0, 0))
zp2 <- zp2 + scale_y_continuous("Proportion", expand = c(0, 0))
zp2 <- zp2 + scale_fill_discrete("Factor Level")
zp2 <- zp2 + theme_bw()
print(zp2)

Holland$class = NULL
Holland$class_media = holland_media_lca2$predclass
Holland$class_media = factor(Holland$class)

#Age
boxplot(Q22_Age~class_media,data=Holland, main="Age by Class",
        xlab="Class", ylab="Age")
pairwise.t.test(Holland$Q22_Age, Holland$class_media, p.adj = "bonferroni")
#test de anova
ajuste2 <- lm(Holland$Q22_Age ~ Holland$class_media)
anova(ajuste2)

#gender
round(prop.table(table(Holland$class_media, Holland$Q21_Sex),1)*100, 1)
pairwise.t.test(Holland$Q21_Sex, Holland$class_media, p.adj = "bonferroni")
CrossTable(Holland$Q21_Sex,Holland$class_media, prop.chisq = F,format = "SPSS",digits = 2,chisq=2)

#Age with 5 categories
CrossTable(Holland$Q22_Age_5cat,Holland$class_media, prop.chisq = F,format = "SPSS",digits =
1,chisq = T)

#Income
CrossTable(Holland$income_3cat,Holland$class_media,prop.chisq = F,format = "SPSS",digits = 1,chisq
= T)

#Family status
he<-Holland[!(Holland$has_partner=="NR"),]
CrossTable(he$has_partner,he$class_media, prop.chisq = F,format = "SPSS",digits = 2,chisq = T)

#Has Children?
hc<-Holland[!(Holland$has_children=="NR"),]
CrossTable(hc$has_children,hc$class_media, prop.chisq = F,format = "SPSS",digits = 2,chisq = T)

#employment
hl<-Holland[!(Holland$employ_3cat=="DK_NR"),]
```

```
CrossTable(hl$employ_3cat,hl$class_media, prop.chisq = F,format = "SPSS",digits = 1,chisq = T)

#Education
round(prop.table(table(Holland$class_media, Holland$edu_3cat),1)*100, 1)
#I decided to remove the Don't knows of the variable edu_3cat
hp<-Holland[!(Holland$edu_3cat=="Don't know"),]
round(prop.table(table(hp$class_media, hp$edu_3cat),1)*100, 1)
CrossTable(hp$edu_3cat,hp$class_media, prop.chisq = F,format = "SPSS",digits = 2,chisq = T)

#################The Netherlands mobile usage##############

varmh<-c("Q7_SMS", "Q7_Photos", "Q7_Web_Browser", "Q7_Web_Apps", "Q7_Inst_mess",
        "Q7_SNS", "Q7_Games", "Q7_Calendar","Q7_Alarm", "Q7_GPS_Mapps",
        "Q7_Down_Apps", "Q7_Voice_calls", "Q7_TV_Video"," Q7_Email")


for (n in varmh){
  print(n)
  print(round(100*prop.table(table(
    holland_mobile_lca3$predclass, na.omit(Holland[[n]])),1), 1))
}

print(round(100*prop.table(table(
  holland_mobile_lca3$predclass, na.omit(Holland[["Q7_Email"]])),1), 1))
for (n in varmh){
  print(n)
  print(CrossTable(
    holland_mobile_lca3$predclass, na.omit(Holland[[n]])))
}
##GG plot
lcModel = holland_mobile_lca3
lcModelProbs <- melt(lcModel$probs)
str(factor(holland_media_lca3$predclass))

zp2 <- ggplot(lcModelProbs,
             aes(x = Var1, y = value, fill = Var2))
zp2 <- zp2 + geom_bar(stat = "identity", position = "stack")
zp2 <- zp2 + facet_wrap(~ L1)
zp2 <- zp2 + scale_x_discrete("Class", expand = c(0, 0))
zp2 <- zp2 + scale_y_continuous("Proportion", expand = c(0, 0))
zp2 <- zp2 + scale_fill_discrete("Factor Level")
zp2 <- zp2 + theme_bw()
print(zp2)

complete <- function(data, desiredCols) {
  completeVec <- complete.cases(data[, desiredCols])
  return(data[completeVec, ])
}


netherlands<-complete(Holland,"Q7_SMS")
netherlands$class = NULL
netherlands$class =holland_mobile_lca3$predclass
netherlands$class = factor(netherlands$class)


#gender
round(prop.table(table(netherlands$class, netherlands$Q21_Sex),1)*100, 1)
pairwise.t.test(netherlands$Q21_Sex,netherlands$class,p.adj="bonferroni")
```

```
CrossTable(netherlands$class, netherlands$Q21_Sex,format = "SPSS",prop.chisq = F,digits = 2,chisq =
T)

#age in 5 cat
round(prop.table(table(netherlands$class, netherlands$Q22_Age_5cat),1)*100, 1)
CrossTable(netherlands$class, netherlands$Q22_Age_5cat,format = "SPSS",prop.chisq = F,digits =
1,chisq = T)
#income
round(prop.table(table(netherlands$class, netherlands$income_3cat),1)*100, 1)
CrossTable( netherlands$income_3cat,netherlands$class,format = "SPSS",prop.chisq = F,digits = 1,chisq
= T)

#Family status
round(prop.table(table(netherlands$class, netherlands$has_partner),1)*100, 1)
np<-netherlands[!(netherlands$has_partner=="NR"),]
CrossTable(np$class, np$has_partner,prop.chisq = F,format = "SPSS",digits = 2,chisq = T)

#has children?
round(prop.table(table(netherlands$class, netherlands$has_children),1)*100, 1)
nc<-netherlands[!(netherlands$has_children=="NR"),]
CrossTable(nc$class, nc$has_children,prop.chisq = F,format = "SPSS",digits = 2,chisq = T)

#Employment
np<-netherlands[!(netherlands$employ_3cat=="DK_NR"),]
CrossTable(np$class, np$employ_3cat,prop.chisq = F,format = "SPSS",digits = 2,chisq = T)


#education
round(prop.table(table(netherlands$class, netherlands$edu_3cat),1)*100, 1)
#I decided to remove the Don't knows of the variable edu_3cat
ne<-netherlands[!(netherlands$edu_3cat=="Don't know"),]
CrossTable(ne$class, ne$edu_3cat,prop.chisq = F,format = "SPSS",digits = 2,chisq = T)


###########Crosstable Media usage vs mobile usage#############

#Canada
CrossTable(svytable(~class_media+class, design = canada.mo),prop.chisq = F,format ="SPSS",chisq=T)

#spain
CrossTable(Spain_mob$class,Spain_mob$class_media,chisq=T,prop.chisq = F,format ="SPSS")

#Netherlands
CrossTable(netherlands$class_media, netherlands$class,chisq=T,prop.chisq = F,digits = 2,format
="SPSS")
```