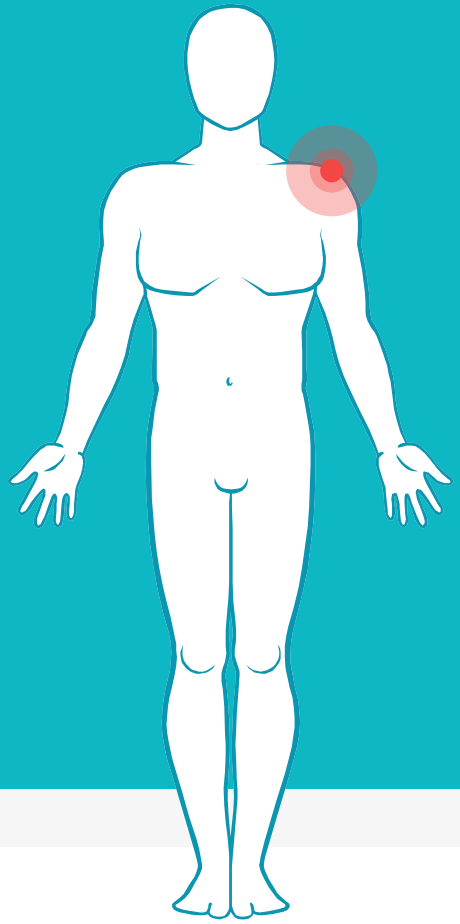
Dietary habits and disease activity in rheumatoid arthritis patients



2

RESEARCH QUESTIONS

RETRIEVE DIETARY PATTERNS

- Are there prevalent dietary patterns (DPs) followed by patients?

TECHNIQUES

- Principal component analysis
- Factor analysis

RELATE PATTERNS WITH DISEASE ACTIVITY

- Are there significant relationships between different dietary patterns and disease form?

TECHNIQUES

- Clustering
- (Logistic Regression)

Analysis supported by

Edefonti V, Parpinel M, Ferraroni M, Boracchi P, Schioppo T, Scotti I, Ubiali T, Currenti W, De Lucia O, Cutolo M, Caporali R, Ingegnoli F. A Posteriori Dietary Patterns and Rheumatoid Arthritis Disease Activity: A Beneficial Role of Vegetable and Animal Unsaturated Fatty Acids. Nutrients. 2020 Dec 17;12(12):3856. doi: 10.3390/nu12123856. PMID: 33348651; PMCID: PMC7766886.

1. DATASET

PRE-PROCESSING

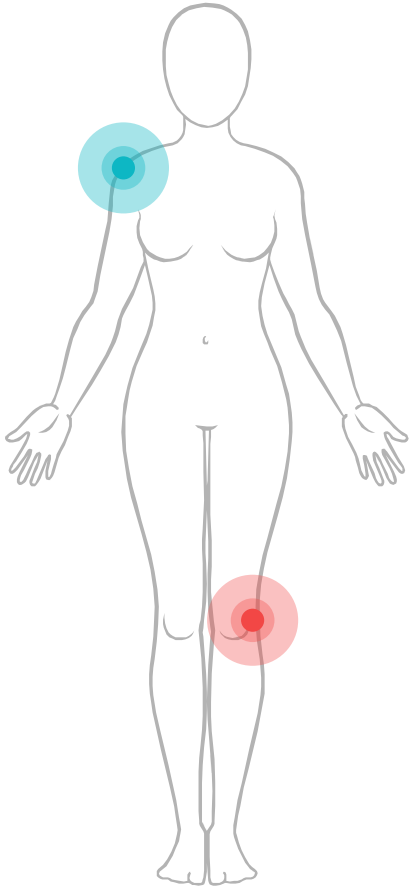
365 observations, 499 variables.

Each row is a patient suffering of rheumatoid arthritis.

RELEVANT ATTRIBUTES

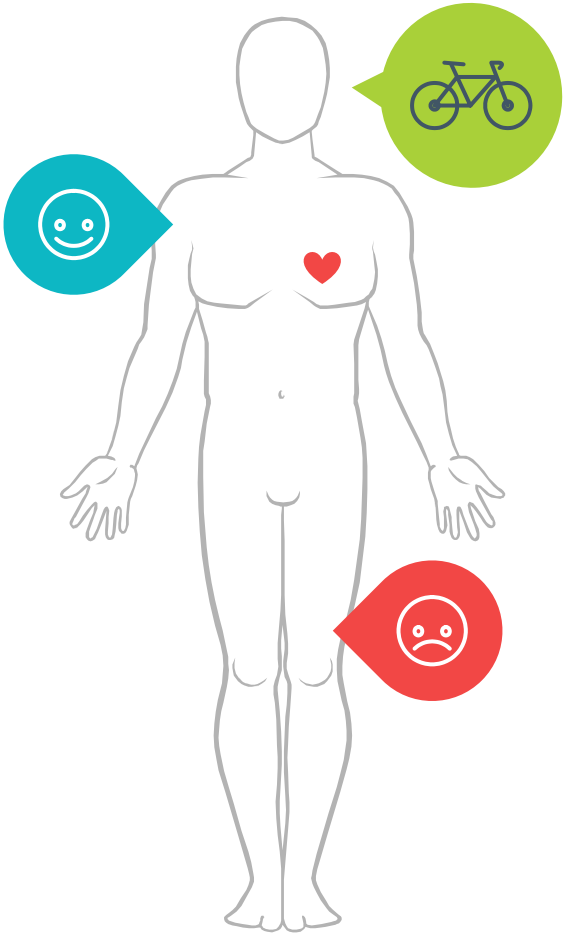
- Patients' info
- Nutrients (34 variables)
- Single foods (110 variables)
- Food groups (35 variables)

- Others: Diseases, drugs, etc.



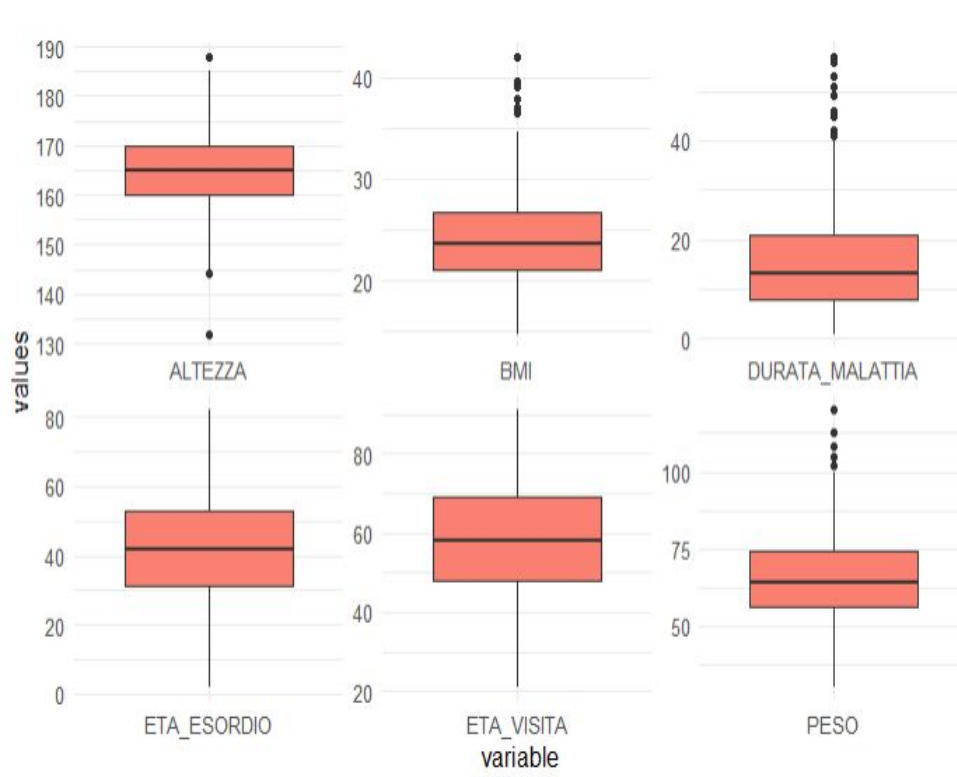
IRRELEVANT ATTRIBUTES

- Attributes for a study on the relation between air pollution and rheumatoid arthritis
- Answers to the food frequency questionnaire: already translated into scores



365 observations, 304 variables left

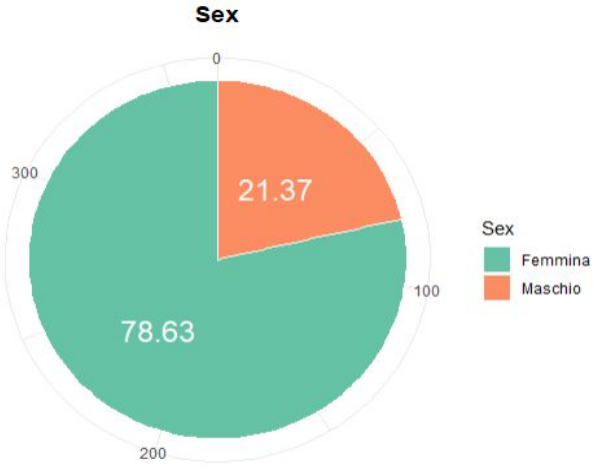
EXPLORATORY ANALYSIS

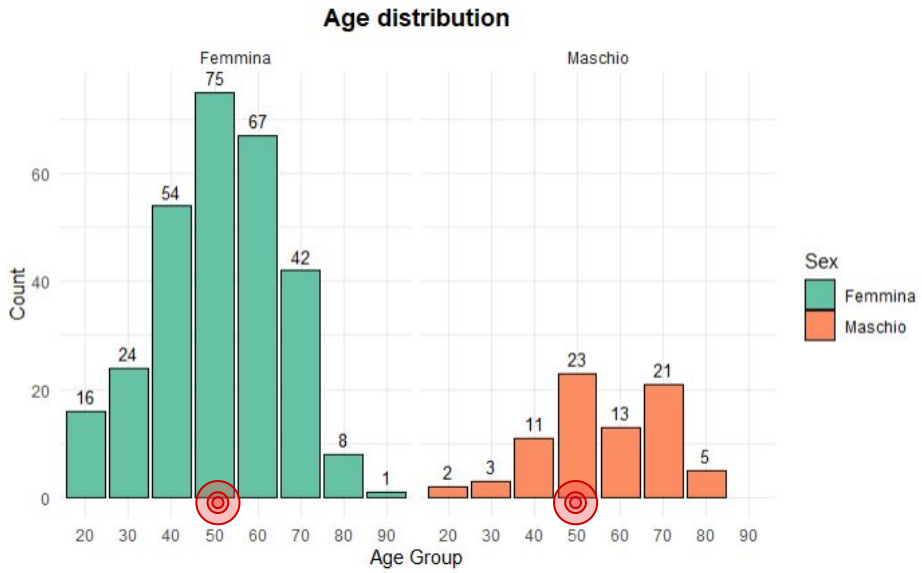


Descriptive analysis of the most relevant variables

BIOLOGICAL SEX and AGE

- Highly unbalanced dataset
- Nearly **80%** of patients are female patients. Disease mainly affecting **female** individuals



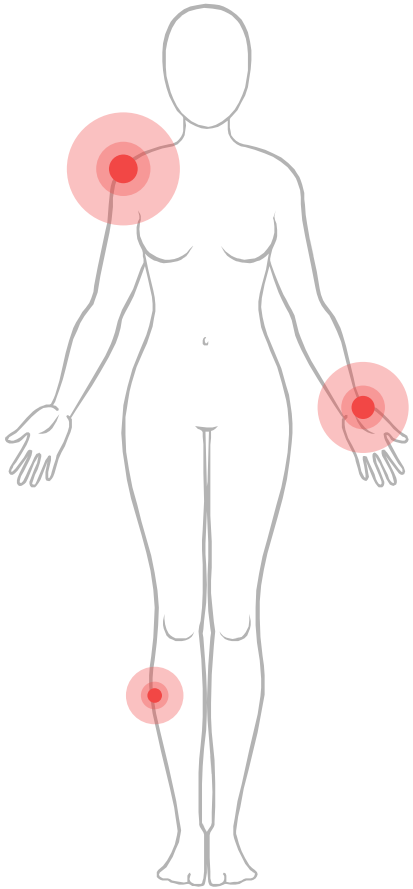


- In both groups most of the patients are in their **50s**

Beginning	Present	Duration

ETA_ESORDIO	ETA_VISITA	DURATA_MALATTIA
Min. : 2.00	Min. :21.00	Min. : 1.00
1st Qu.:31.00	1st Qu.:48.00	1st Qu.: 8.00
Median :42.00	Median :58.00	Median :13.00
Mean :41.88	Mean :57.51	Mean :15.72
3rd Qu.:53.00	3rd Qu.:69.00	3rd Qu.:21.00
Max. :82.00	Max. :91.00	Max. :57.00

Since the average duration is 15 years, the highest **incidence** is around the **30s**



DISEASE ACTIVITY INDICATORS

$$DAS28 = (0.56 * sqr(CAD)) + (0.28 * sqr(CAG)) + (0.7 * ln(VES)) + (0.014 * GH)$$

RHEUMATOID FACTOR

{0,1}

- 1: severe activity
- 0: low-moderate activity

DAS28

(0.0,9.9)

- **DAS28 < 2,6:** Remission
- **DAS28 >= 2,6 & <= 3,2:** Low activity
- **DAS28 > 3,2 & <= 5,1:** Moderate activity
- **DAS28 > 5,1:** Severe activity

Composite index

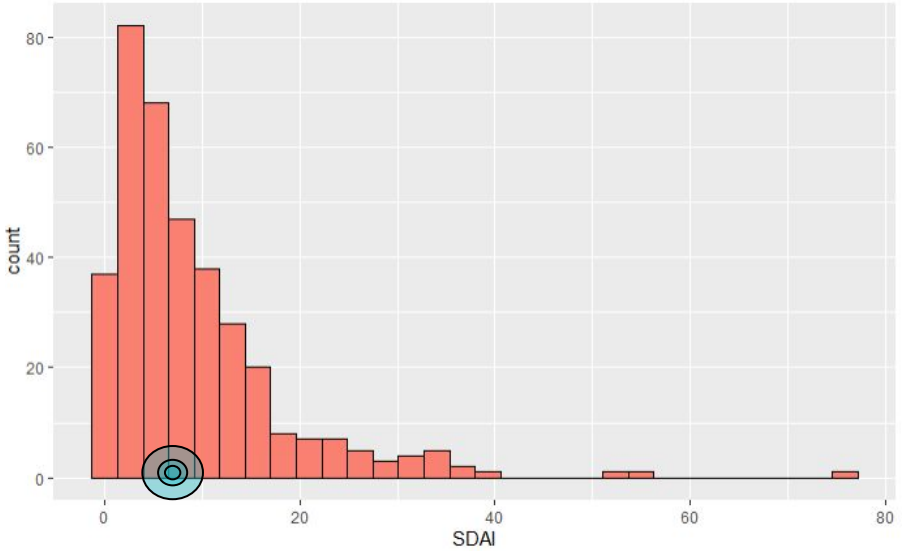
SDAI

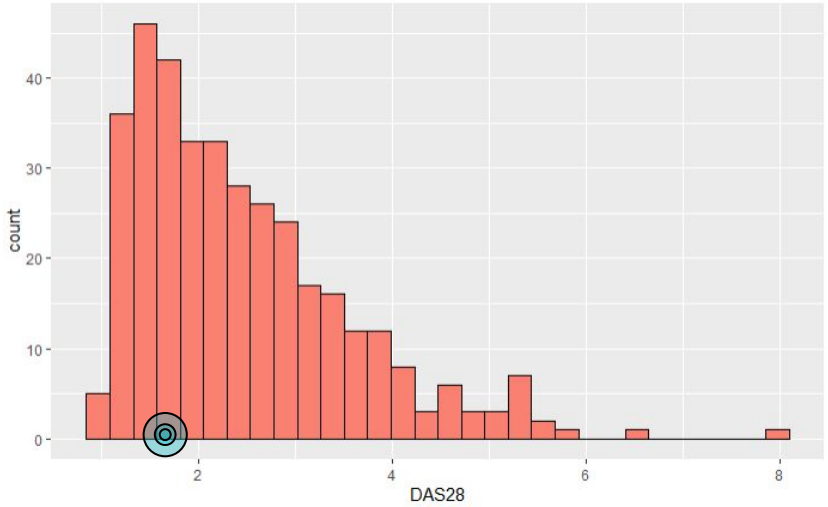
(0.0,100.0)

- **SDAI <3.3:** Remission
- **SDAI >= 3.3 & <= 11:** Low activity
- **SDAI > 11 & <= 26:** Moderate activity
- **SDAI > 26:** Severe activity

Composite index (Simplified Disease Activity Index): measuring pain in 28 joints


0
100
200
300
46.3
53.7
FR
0
1


count
SDAI
0
20
40
60
80


count
DAS28
0
10
20
30
40
2
4
6
8

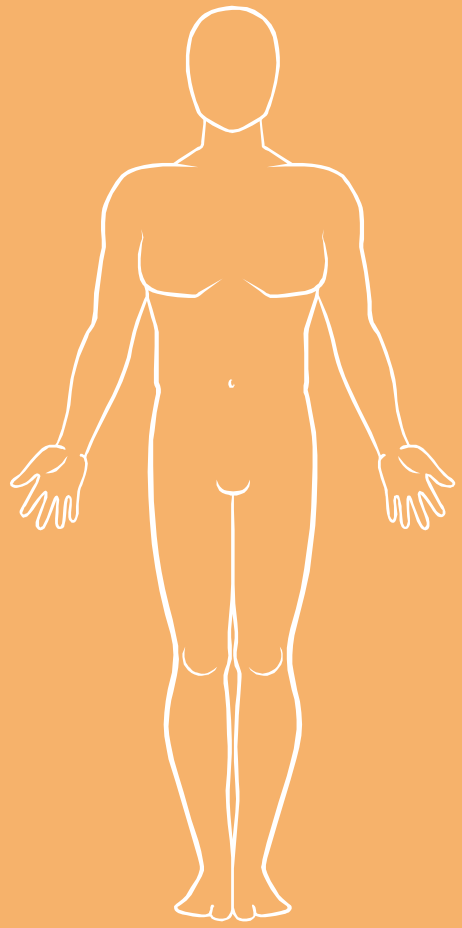
SDAI

108
REMISSION

155
LOW ACTIVITY

81
MODERATE ACTIVITY

21
SEVERE ACTIVITY



DAS28

227
REMISSION

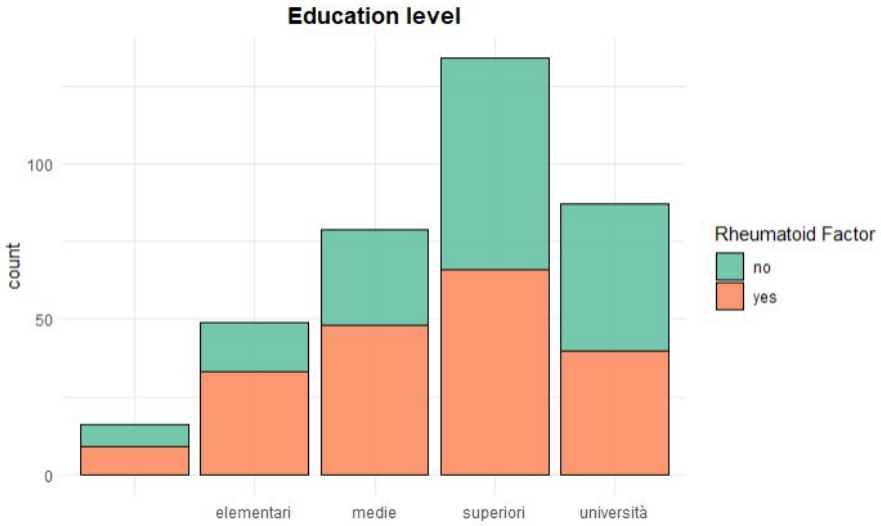
60
LOW ACTIVITY

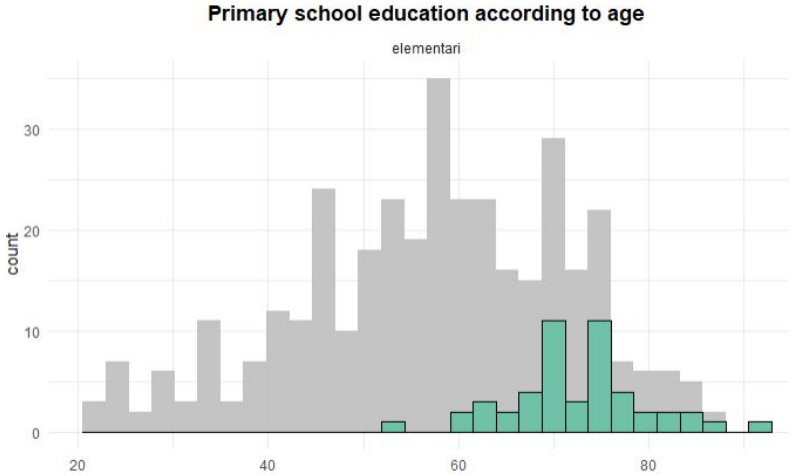
64
MODERATE ACTIVITY

14
SEVERE ACTIVITY

EDUCATION

- 16 missing values
- No expected relationship between disease activity and education level

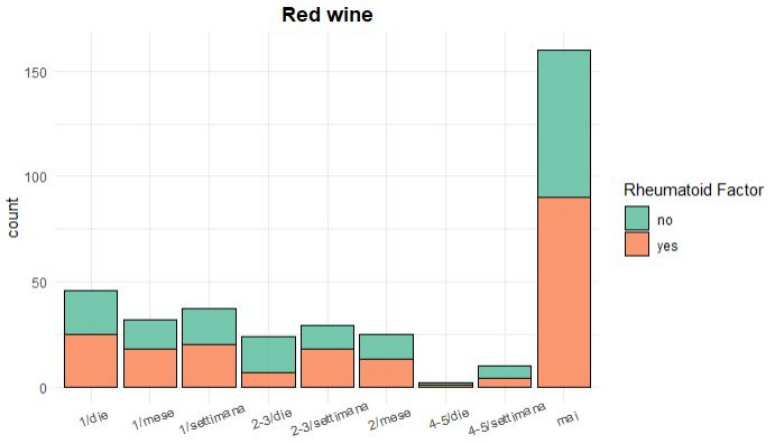


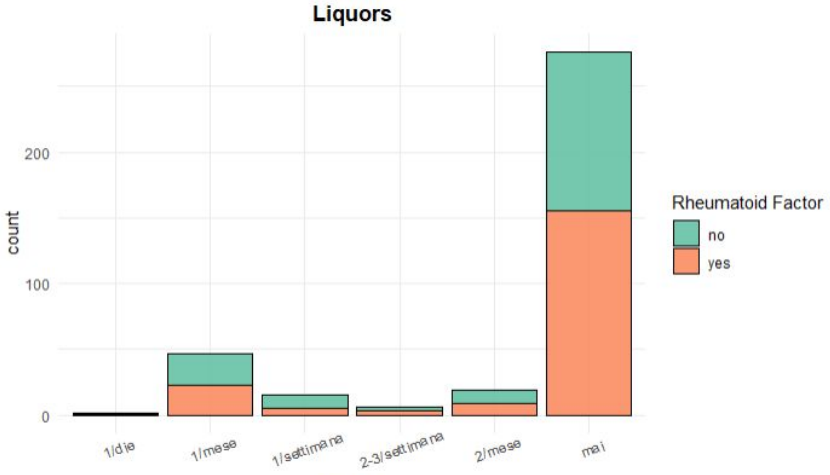


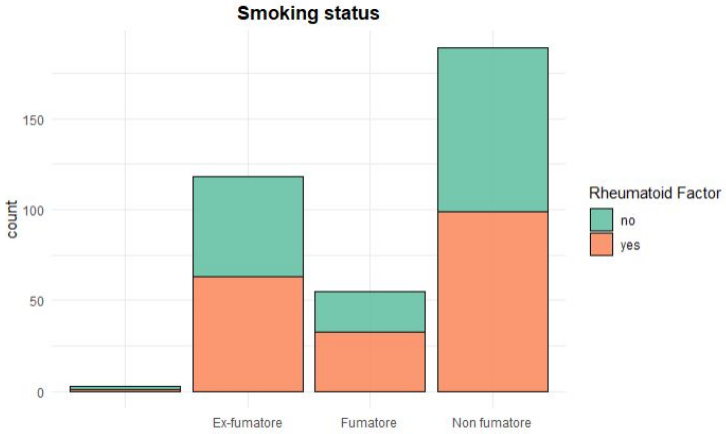
Higher incidence of 1 in primary school due to the age of patients: older people more likely to have the disease and more likely to have finished primary school only. **No correlation**

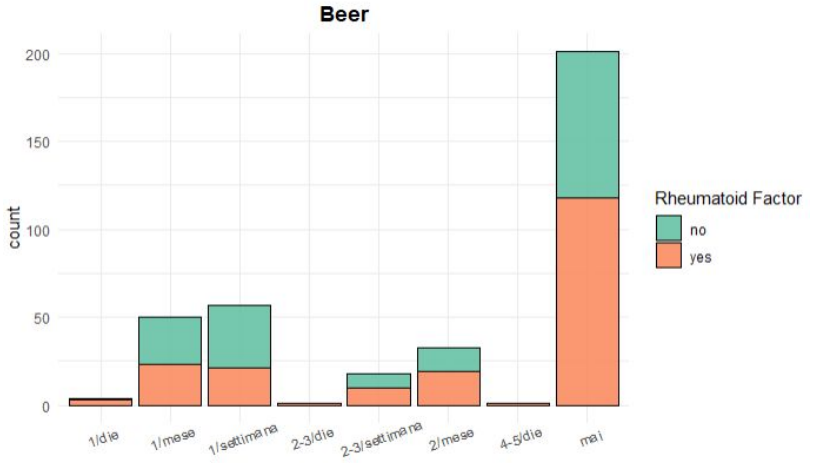
ALCOHOL and SMOKE

No relevant relationship

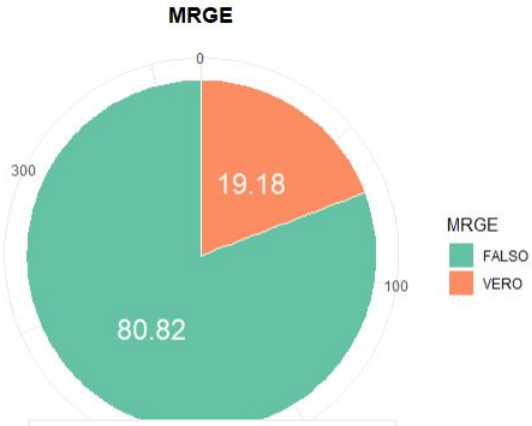




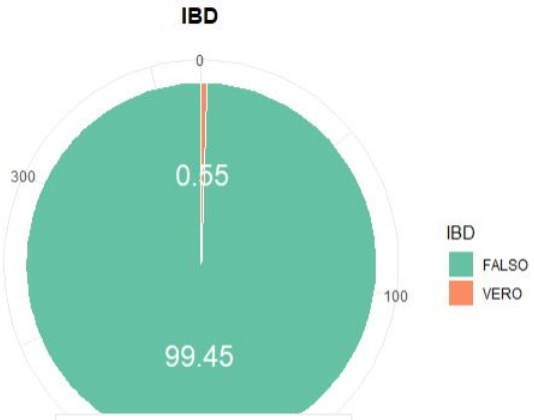




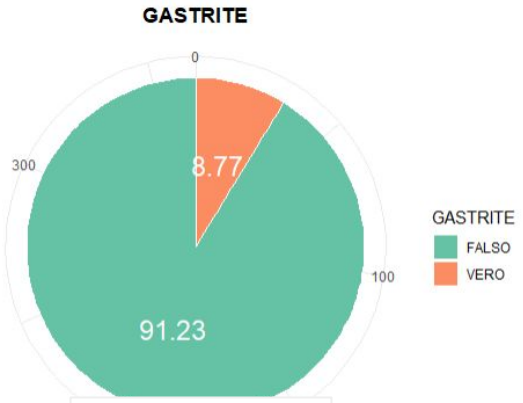
COMORBIDITIES



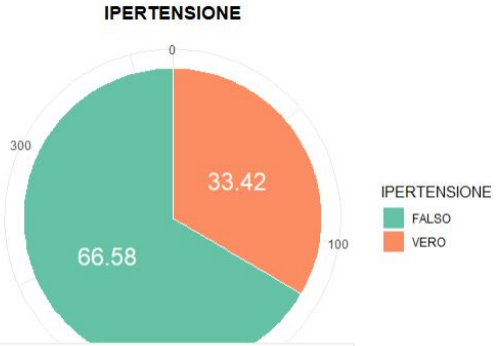
Gastroesophageal Reflux



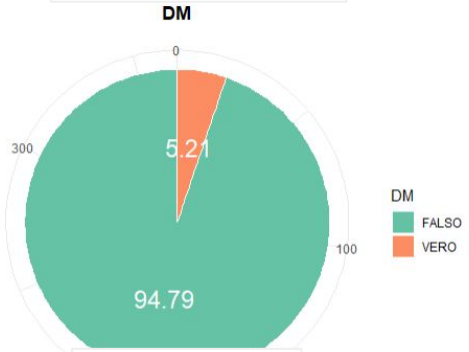
Inflammatory bowel



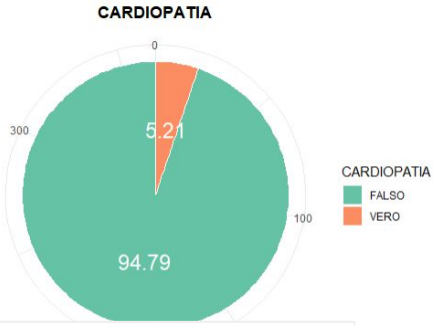
Gastritis



Hypertension



Diabetes



Cardiovascular

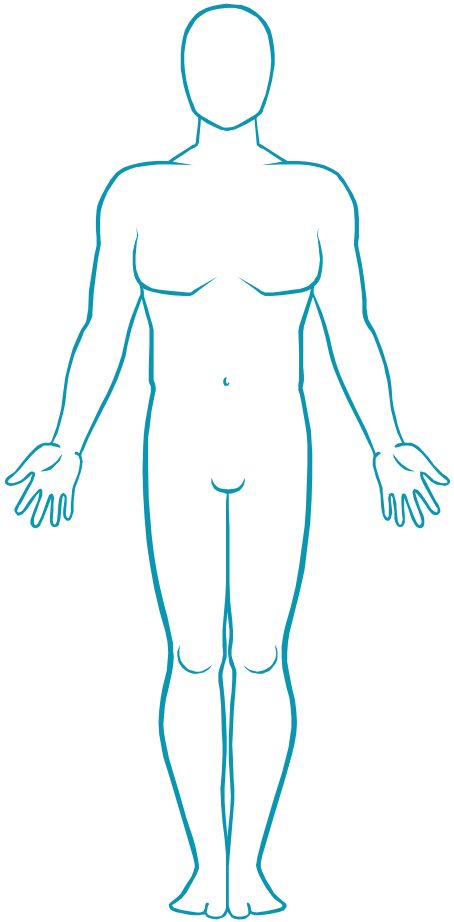
2.
DPs RETRIEVAL

PCA and FACTOR ANALYSIS

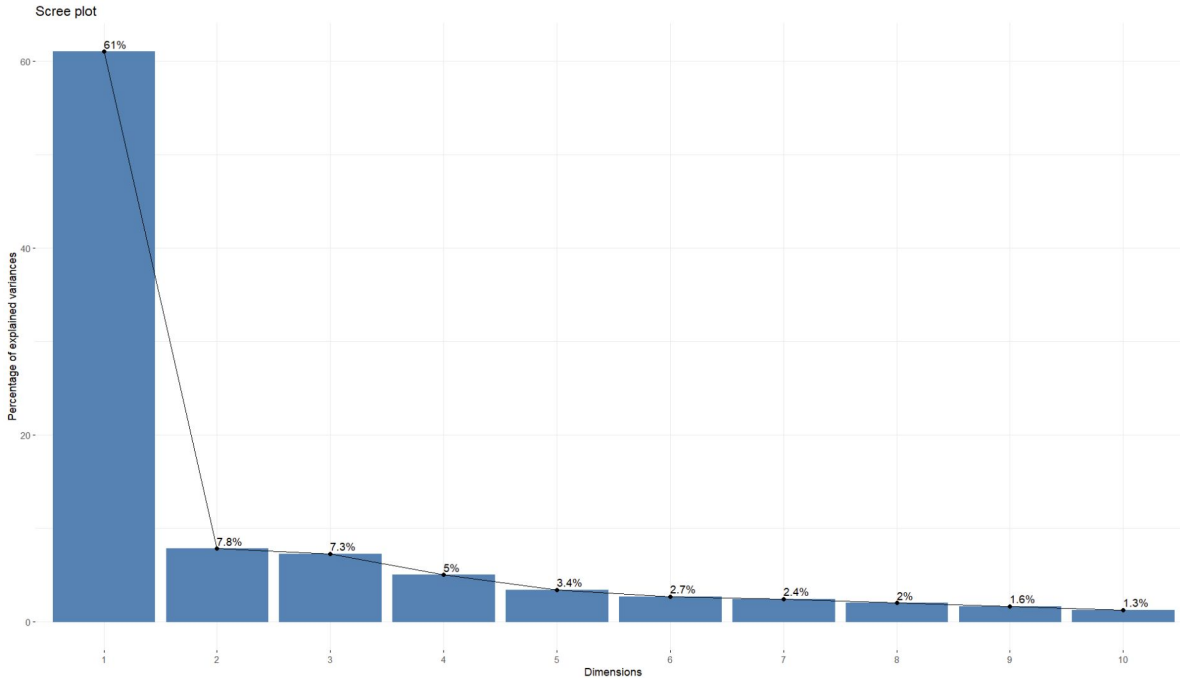
Principal Component Analysis

- Goal: Reduce dataset and find suitable number of components explaining the majority of the variance in the dataset
- PCA was performed on the original dataset as well as subsets, split based on the DAS28 score to find possible differences between lightly/heavily affected patients

- Only subset with high DAS28 (>3.2) Included in presentation

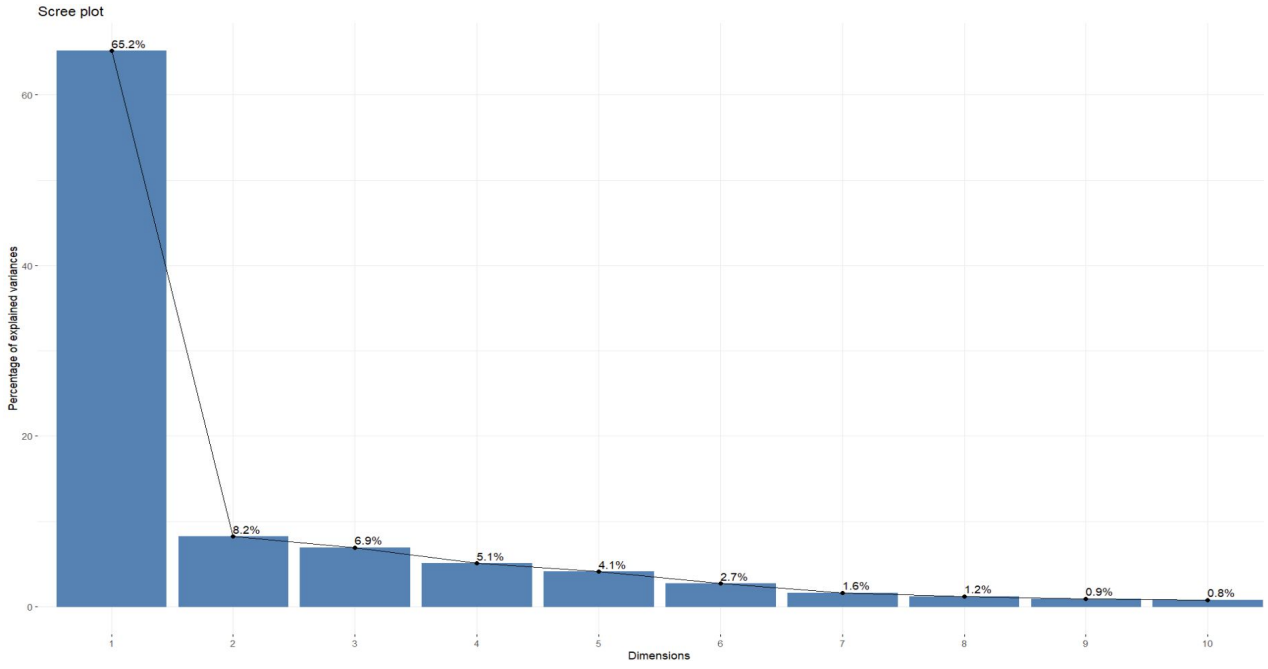


Principal Component Analysis



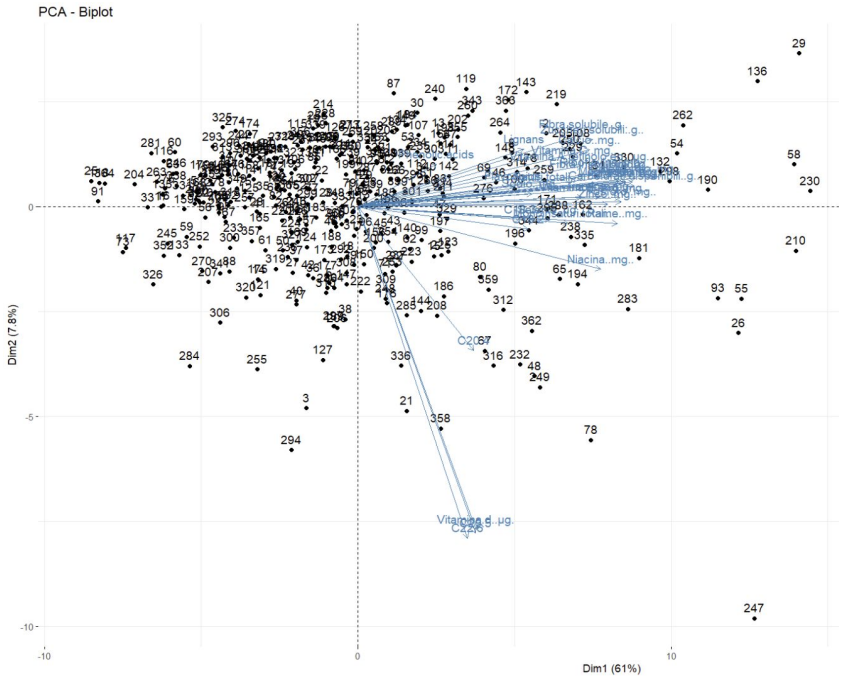
	Cumulative Proportion
PC1	0,610
PC2	0,688
PC3	0,761
PC4	0,811
PC5	0,845
PC6	0,872

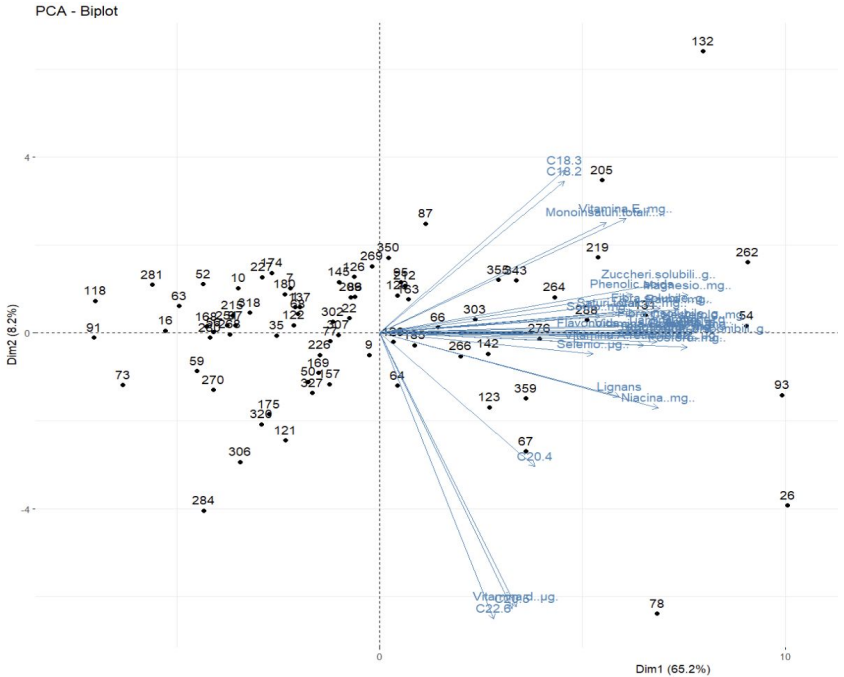
Principal Component Analysis – high DAS28 score



	Cumulative Proportion
PC1	0,6515
PC2	0,7337
PC3	0,8031
PC4	0,8539
PC5	0,8947
PC6	0,9218

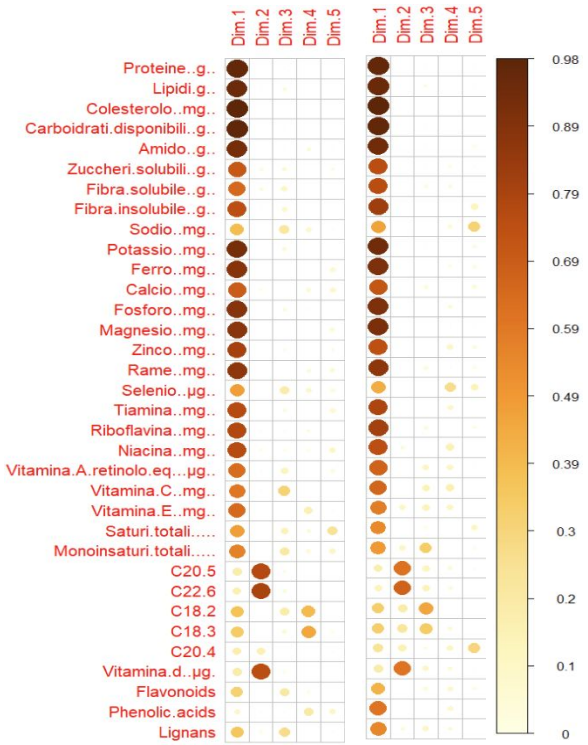
Principal Component Analysis





Principal Component Analysis

- Table of PC1-5 including the full set of individuals
- To find the most contributing variables and derive insights about the nature of the components and their driving factors for the individual components we checked the correlations
- C20.5, C22.6 are omega3 fatty acids
- C20.4 (arachidonic),
- C18.2 (linoleic), C18.3 (linolenic) omega6



Factor Analysis - loadings

- Goal: retrieve 5 factor loadings for the components found in the PCA and associate them to possible dietary patterns
- FA has been performed on set of 34 nutrients included in the dataset
- Outstanding aspects in terms of loading for the nutrients have been highlighted
- 5 DP have been formulated

Loadings:

	Factor1	Factor2	Factor3	Factor4	Factor5
Proteine..g..	0.844				0.155
Lipidi.g..			0.638		0.601
Colesterolo..mg..	0.180		-0.145	0.172	0.774
Carboidrati.disponibili..g..	0.816			-0.145	
Amido..g..	0.866	-0.325		-0.145	
Zuccheri.solubili..g..		0.800			0.212
Fibra.solubile..g..	0.182	0.806			
Fibra.insolubile..g..	0.318	0.711			-0.125
Sodio..mg..	0.545	-0.147			0.329
Potassio..mg..	0.294	0.744			
Ferro..mg..	0.729	0.363			-0.142
Calcio..mg..	0.251	0.432	-0.126		0.533
Fosforo..mg..	0.698	0.217			0.181
Magnesio..mg..	0.771	0.306	0.131		-0.172
Zinco..mg..	0.714	0.243			
Rame..mg..	0.659	0.278	0.283		-0.224
Selenio..µg..	1.059	-0.251			
Tiamina..mg..	0.728	0.194	0.100		
Riboflavina..mg..	0.448	0.390			0.245
Niacina..mg..	0.904			0.166	-0.137
Vitamina.A.retinolo.eq...µg..		0.865			0.230
Vitamina.C..mg..	-0.142	1.065			
Vitamina.E..mg..		0.356	0.683		
Saturi.totali.....			0.179		0.923
Monoinsaturi.totali.....		-0.108	0.714		0.477
C20.5				0.964	
C22.6				1.000	
C18.2		-0.179	0.966		
C18.3			0.870		
C20.4	0.332			0.273	0.150
Vitamina.d..µg.				0.956	
Flavonoids		0.717			
Phenolic.acids		0.280	0.206		-0.189
Lignans	-0.123	0.877	-0.173		

Factor Analysis - Correlations

- The factor loadings have then been analysed in a correlation matrix with the food groups from the dataset to check validity of the dietary patterns
- They show the strongest correlations with the groups we would expect from each dietary pattern

	Starch_rich	Vitamins_fiber	VUFA	AUFA	Animal_prod
vf_fast_food	0.25	-	-	-	-
vf_formaggio_spalmabile	0.2	0.17	-	-	0.41
vf_redmeat	0.34	0.16	-	0.13	0.34
vf_whitemeat	-	-	-	-	0.17
vf_fish	0.19	0.25	0.15	0.72	0.11
vf_fat_fish	0.23	0.20	0.20	0.73	0.19
vf_seafood	-	0.13	-	0.16	-
vf_eggs		-	-	-	-
vf_sweets	0.24	0.13	0.29	-	0.42
vf_snacks	0.17	-	0.24	-	0.26
vf_dried_fruit		-	0.17	-	-
vf_nuts	0.26	0.24	0.85	0.12	-
vf_olive_oil	0.18	0.22	0.37	0.18	-
vf_seed_oil		-	-	-	-
vf_vegetables	0.45	0.73	0.35	0.28	0.24
vf_leafy_vegetables	0.37	0.60	0.33	0.27	0.19
vf_cabbages	0.29	0.55	0.19	0.19	0.13
vf_legumes	0.57	0.44	0.32	0.24	0.16
vf_onion_garlic	0.29	0.31	0.21	0.21	0.11
vf_mushrooms	0.29	0.27	0.15	0.22	0.23
vf_soy_products	0.15	0.15	0.21	-	-
vf_fruits_allseasons	0.37	0.79	0.38	0.22	0.24
vf_winter_fruits	0.31	0.73	0.20	0.18	0.24
vf_summer_fruits	0.34	0.74	0.37	0.17	0.22
vf_unsweetened_beverages	0.15	0.31	0.18	0.17	-
vf_coffee	0.19	0.14	-	-	0.14
vf_alcoholic_beverages	-	-	-	-	-
vf_soft_drinks	-	-	-	-	0.12
vf_refined_grains	0.55	-	0.2	-	0.33
vf_potatoes	0.31	0.19	0.21	0.16	0.29
vf_whole_grains	0.64	0.24	0.3	-	0.13
vf_refined_grains	0.55	-	0.2	-	0.33
vf_potatoes	0.31	0.19	0.21	0.16	0.29
vf_whole_grains	0.64	0.24	0.3	-	0.13
vf_butter_margarine	0.11	-	0.11	-	0.35
vf_milk_yoghurt	0.25	0.18	-	-	0.37
vf_stagionato_cheese	0.31	0.35	0.17	-	0.67
vf_other_cheese	0.37	0.25	0.11	0.12	0.65

Adherence to DP's

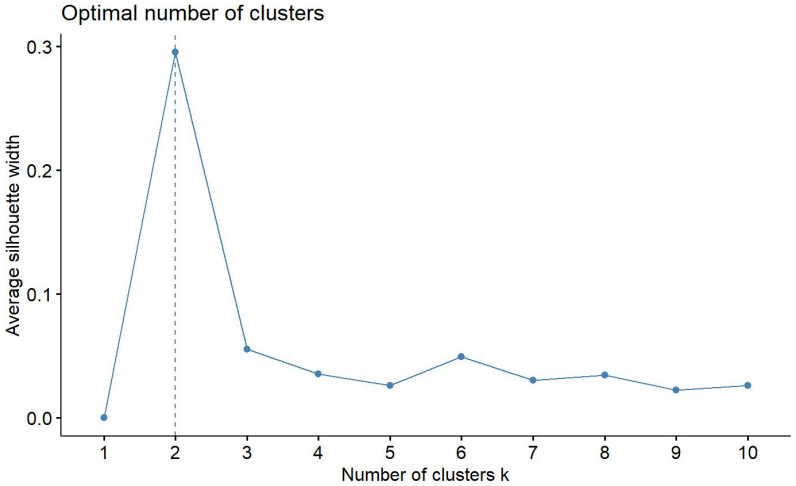
- We assigned the DP's to the individuals according to the highest factor loading

Stark-rich	Vitamins_ Fiber	VUFA	AUFA	Animal Products
53	65	49	101	97

Factor1 <dbl>	Factor2 <dbl>	Factor3 <dbl>	Factor4 <dbl>	Factor5 <dbl>	max <dbl>	Pattern <dbl>
-0.089557962	-0.952617282	-0.309470804	-0.699896011	-0.432295498	-0.089557962	1
0.176391754	2.896447533	0.095706199	0.072934347	-0.475993577	2.896447533	2
-0.393344158	-0.496801613	0.434579857	2.618567038	-0.458250869	2.618567038	4
-0.214375554	-0.592607424	0.062558390	-0.124279305	-0.145708982	0.062558390	3
-0.345842414	0.038098349	-0.700194240	1.197109111	-0.315964828	1.197109111	4
0.020924569	-0.236465336	0.019487269	-0.445499120	1.622782675	1.622782675	5
-0.417339508	-0.551359920	-0.817745857	-0.775288183	0.962940720	0.962940720	5
0.370064100	1.718206606	0.367929245	-0.199041084	0.176111789	1.718206606	2
-0.121462369	-0.043110154	-0.104176233	0.118788140	0.723633343	0.723633343	5
-0.633622025	-0.917078655	-0.517700190	-0.996676432	-1.030894846	-0.517700190	3

3.
CLUSTERING

Confirmatory analysis



- Despite Silhouette method result, 5 clusters were identified using k-means

- Attempt with 5 clusters revealed an outlier
- K-means with 6 clusters to have 5 actual groups

```
> table(km.outcome$cluster)

  1   2   3   4   5   6
103  93  46  15   1 107
```

	ID <int>	SESSO <fctr>	ALTEZZA <int>	PESO <int>	BMI <dbl>	SCOLORITA <fctr>	CITTADINANZA <chr>	Fumatore <fctr>	FR <fctr>
272	300	Femmina	168	55	19.48696	medie	italiana	Ex-fumatore	1

- Average consumption scores for different clusters

	vf_carni_rosse <dbl>	vf_maiale <chr>	vf_fast food <chr>	vf_carni_bianca <dbl>	vf_carni_processata <dbl>
1	0.16	-	-	0.30	0.17
2	0.19	-	-	0.24	0.25
3	0.23	0.08	0.01	0.28	0.24
4	0.25	0.06	0.03	0.28	0.25
→ 5	0.36	0.36	0	0.64	0.64
6	0.20	0.09	0.01	0.25	0.25

	vf_pere <dbl>	vf_arance_mandarini <dbl>	vf_arance_rosse <dbl>	vf_banana <dbl>	vf_uva <dbl>	vf_melone <dbl>	vf_anguria <dbl>	vf_pesche <dbl>	vf_albicocche <dbl>
1	0.20	0.51	0.35	0.28	0.20	0.20	0.17	0.37	0.35
2	0.23	0.59	0.45	0.33	0.30	0.30	0.29	0.55	0.50
3	0.66	1.83	0.87	0.56	1.02	0.48	0.53	1.47	1.67
4	1.48	2.19	1.61	0.79	1.69	1.19	1.08	2.10	2.78
→ 5	2.50	4.50	4.50	0.00	0.00	0.36	4.50	4.50	4.50
6	0.28	0.73	0.39	0.25	0.25	0.25	0.29	0.51	0.53

	vf_spinaci <dbl>	vf_broccoli <chr>	vf_cavoli <chr>	vf_cavolfiori <chr>	vf_piselli <dbl>	vf_fagiolini <dbl>	vf_zucchine <dbl>	vf_cicoria <chr>	vf_asparagi <chr>	vf_insalata <dbl>
1	0.13	0.13	-	-	0.11	0.13	0.23	0.12	-	0.44
2	0.11	-	-	-	0.11	0.12	0.22	-	-	0.42
3	0.14	0.19	0.12	0.16	0.17	0.26	0.27	0.2	0.13	0.96
4	0.29	0.29	0.17	0.23	0.26	0.37	0.36	0.13	0.17	1.39
→ 5	4.50	4.5	4.5	4.5	0.36	0.14	0.64	0.06	4.5	0.36
6	0.12	0.13	0.08	0.1	0.10	0.14	0.24	0.1	0.09	0.63

- Hard to find one-to-one correspondence with DPs (e.g. 4th cluster could be **Vitamins and fiber** DP)

	vf_pane_farinacei <dbl>	vf_pane_farinacei_integrali <dbl>	vf_cornflakes_cereali <chr>	vf_riso <dbl>	vf_pasta <dbl>	vf_riso_pasta_integrali <chr>	vf_patate <dbl>
1	0.67	0.32	0.14	0.24	0.43	-	0.15
2	1.28	0.47	-	0.29	0.62	0.12	0.15
3	1.26	0.55	0.18	0.29	0.56	0.16	0.19
4	0.80	0.87	0.17	0.27	0.31	0.15	0.19
5	0.64	0.06	0	0.14	0.14	0	0.03
6	1.26	0.60	0.17	0.22	0.47	0.14	0.14

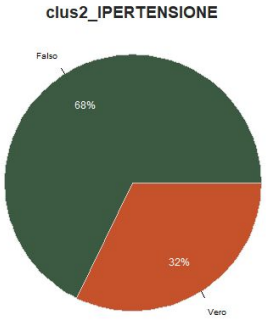
	vf_formaggio_giallo <dbl>	vf_ricotta_formaggi_magri <dbl>	vf_mozzarella <dbl>	vf_formaggio spalmabile <chr>	vf_burro <dbl>	vf_margarina <chr>	vf_dolci_vari <dbl>
1	0.25	0.18	0.11	-	0.14	-	0.53
2	0.38	0.20	0.18	0.12	0.25	-	0.85
3	0.59	0.32	0.21	0.22	0.20	0.02	1.45
4	0.51	0.29	0.18	0.15	0.12	0.03	0.99
5	2.50	0.06	0.06	0.03	0.00	0.03	0.06
6	0.32	0.26	0.16	0.08	0.12	0.02	0.62

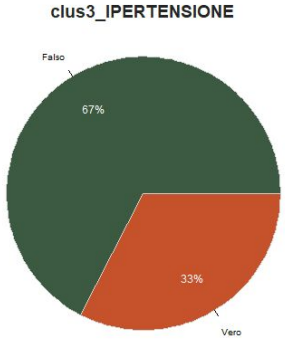
- 2nd and 3rd clusters could be **starch-rich** and **AUFA** respectively

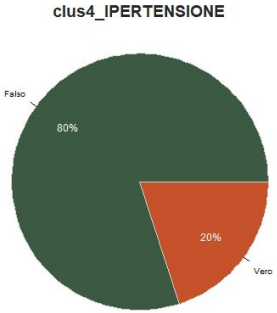
cluster <int>	DAS28 <dbl>	SDAI <dbl>	duration <dbl>
1	2.671083	10.789311	14.17476
2	2.237788	7.194032	15.68817
3	2.587020	10.136457	17.54348
4	1.886081	3.617400	15.80000
5	3.813994	14.080000	16.00000
6	2.478682	9.125897	16.42991

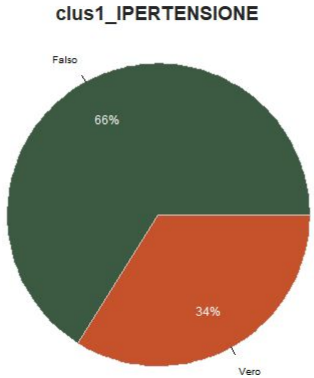
Different comorbidities ratios in the clusters by using most relevant diseases

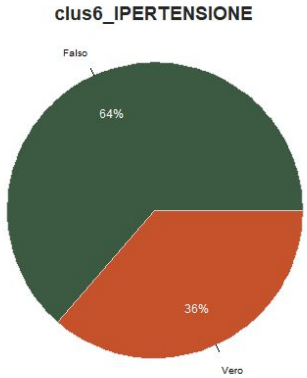
High blood pressure (hypertension)





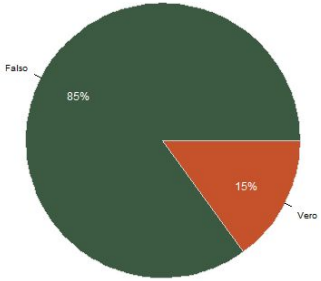




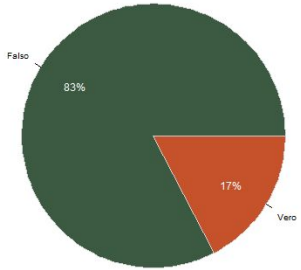


GASTROESOPHAGEAL REFLUX DISEASE (GERD)

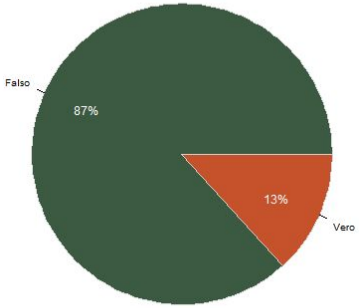
clus2_MRGE



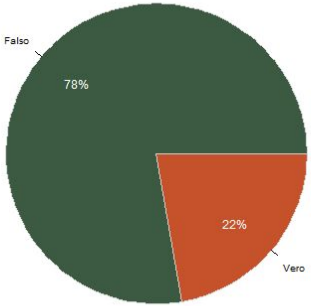
clus3_MRGE



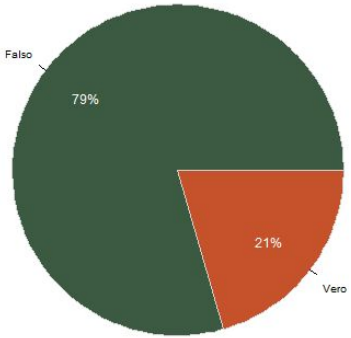
clus4_MRGE



clus1_MRGE

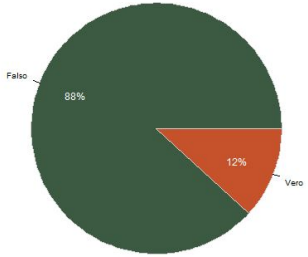


clus6_MRGE

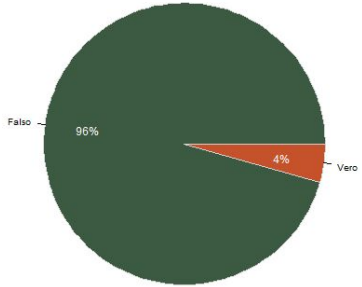


GASTRITIS

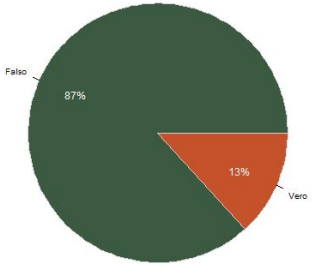
clus2_GASTRITE



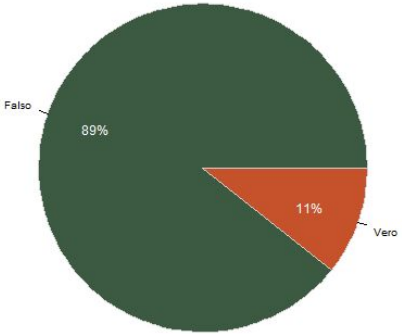
clus3_GASTRITE



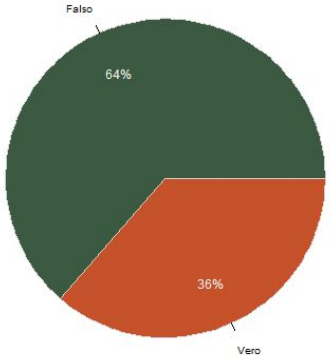
clus4_GASTRITE



clus1_GASTRITE

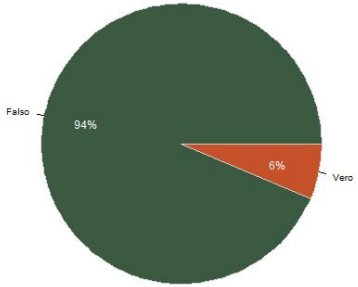


clus6_GASTRITE

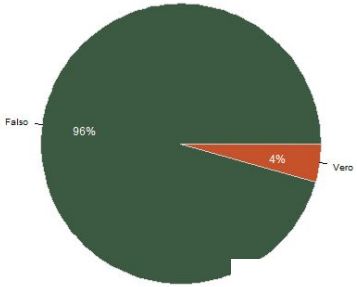


CARDIOVASCULAR

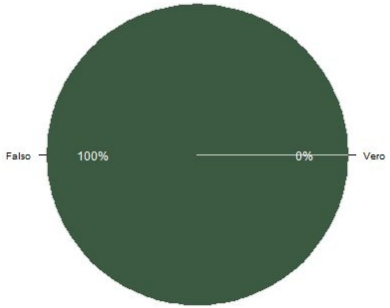
clus2_CARDIOPATIA



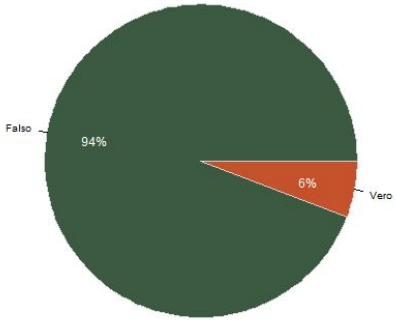
clus3_CARDIOPATIA



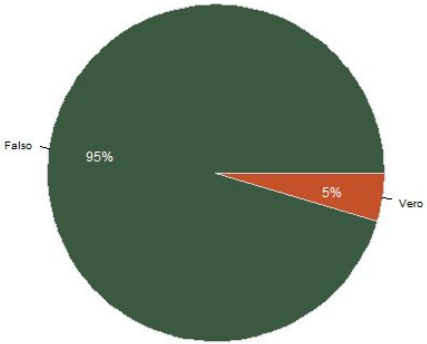
clus4_CARDIOPATIA



clus1_CARDIOPATIA



clus6_CARDIOPATIA



THANK YOU FOR YOUR ATTENTION!

Any questions?