

# Reporting

Descriptive analysis and basic statistics in biomedical studies  
using R and Markdown

Juan R Gonzalez  
juanr.gonzalez@isglobal.org

BRGE - Bioinformatics Research Group in Epidemiology  
ISGlobal - Barcelona Institute for Global Health  
<http://brge.isglobal.org>

IACS - Instituto Aragonés de Ciencias de la Salud  
Zaragoza, February 26th

R facilitates data description and reproducible research

► Patients characteristics comparison

Characteristics	Cases ( <i>n</i> = 416) %	Controls ( <i>n</i> = 1156) %	Value of <i>p</i> <sup>a</sup>
Han race	97.58	96.39	0.25
Education			
None/elementary/high school	71.57	69.24	
Professional/college+	28.43	30.76	0.38
Occupation status			
Physical work	55.42	50.18	
Mental work	44.58	49.82	0.07
BMI [kg/m <sup>2</sup> ]			
Mean ± SD	23.77 ±3.60	23.21 ±2.93	0.01
≥ 24	41.71	34.93	0.05

Figure 1: Baseline comparison table of a standard case-control study

► Odds ratio estimation

Food group	Odds ratio (95% CI) for		
	Colon cancer	Rectal cancer	Colon and rectal cancers
Refined grain	1.46(1.20–1.78)	1.21(0.99–1.49)	1.32(1.12–1.56)
Whole grain	0.92(0.80–1.07)	0.86(0.72–1.02)	0.85(0.75–0.97)
Red meat	1.63(1.30–2.04)	1.50(1.20–1.88)	1.54(1.28–1.85)
Pork and processed meat	1.34(1.17–1.53)	1.18(1.02–1.37)	1.27(1.13–1.43)
Cheese	1.10(0.99–1.22)	1.07(0.94–1.21)	1.09(0.98–1.22)
Raw vegetables	0.90(0.76–1.07)	0.84(0.69–1.01)	0.85(0.74–0.98)
Cooked vegetables	0.69(0.54–0.88)	0.78(0.61–0.99)	0.69(0.57–0.83)
Citrus fruit	0.90(0.79–1.03)	0.84(0.72–0.98)	0.86(0.78–0.96)
Other fruits	0.84(0.71–0.99)	0.87(0.74–1.03)	0.85(0.75–0.96)
Alcohol	1.22(1.04–1.43)	1.38(1.16–1.63)	1.28(1.11–1.48)
Coffee	0.71(0.55–0.92)	0.79(0.62–1.00)	0.73(0.60–0.88)

\*Adjusted for age, sex, education, smoking, alcohol, body mass index, physical activity and total energy intake.

Figure 2: Odds ratio summarization table of a standard case-control study

# compareGroups

compareGroups is an R package available on CRAN to create descriptive tables

It consists of three key functions:

1. `compareGroups`~ generates all the calculation
2. `createTable`~ creates the descriptive table obtained by `compareGroups`. You can costumize it by excluding categories, number of decimals, etc.
3. `export2...`~ exports the descriptive table obtained by `createTable` to EXCEL, Word, LaTeX, Rmarkdown, etc.

# Example

## PREDIMED project:

<http://www.cat.isciii.es/ISCI/III/es/contenidos/fd-el-instituto/fd-comunicacion/fd-noticias/PREDIMED-2013.pdf>

1. Load the package and the example data existing in compareGroups package

```
library(compareGroups)
data(predimed)
# ?predimed
```

```
head(predimed)
```

	group	sex	age	smoke	bmi	waist	wth	htn	diab	hyperchol
1	Control	Male	58	Former	33.53	122	0.7530864	No	No	Yes
2	Control	Male	77	Current	31.05	119	0.7300614	Yes	Yes	No
4	MedDiet + V00	Female	72	Former	30.86	106	0.6543210	No	Yes	No
5	MedDiet + Nuts	Male	71	Former	27.68	118	0.6941177	Yes	No	Yes
6	MedDiet + V00	Female	79	Never	35.94	129	0.8062500	Yes	No	Yes
8	Control	Male	63	Former	41.66	143	0.8033708	Yes	Yes	Yes

	famhist	hormo	p14	toevent	event
1	No	No	10	5.374401	Yes
2	No	No	10	6.097194	No
4	Yes	No	8	5.946612	No
5	No	No	8	2.907598	Yes

## 2. Compute descriptives and other figures by treatment group

- ▶ Use of formula environment to select variables.
- ▶ On left hand side write the variable indicating groups (nothing indicates that descriptive analyses will be performed for the whole database).
- ▶ On the right side write all the variables you want to describe by the grouping variable

```
descr <- compareGroups(group ~ sex + age + smoke, predimed,  
                        method=1) # method = 1 default  
descr
```

----- Summary of results by groups of 'Intervention group'-----

	var	N	p.value	method	selection
1	Sex	6324	<0.001**	categorical	ALL
2	Age	6324	0.003**	continuous normal	ALL
3	Smoking	6324	0.444	categorical	ALL

-----  
Signif. codes: 0 '\*\*' 0.05 '\*' 0.1 ' ' 1

- ▶ method = 1 forces analysis as “normal-distributed”;
- ▶ method = 2 forces analysis as “continuous non-normal”;
- ▶ method = 3 forces analysis as “categorical”; and
- ▶ method = NA, which performs a Shapiro-Wilks test (similar to `ks.test`) to decide between normal or non-normal.

```
descr <- compareGroups(group ~ sex + age + smoke, predimed,
                        method=NA)
descr
```

----- Summary of results by groups of 'Intervention group'-----

	var	N	p.value	method	selection
1	Sex	6324	<0.001**	categorical	ALL
2	Age	6324	0.003**	continuous non-normal	ALL
3	Smoking	6324	0.444	categorical	ALL

-----

Signif. codes: 0 '\*\*\*' 0.05 '\*' 0.1 ' ' 1



- If you are interested in describing all variables use `'`

```
descr <- compareGroups(group ~ ., predimed)
descr
```

----- Summary of results by groups of 'Intervention group'-----

	var	N	p.value	method	selection
1	Sex	6324	<0.001**	categorical	ALL
2	Age	6324	0.003**	continuous normal	ALL
3	Smoking	6324	0.444	categorical	ALL
4	Body mass index	6324	<0.001**	continuous normal	ALL
5	Waist circumference	6324	0.045**	continuous normal	ALL
6	Waist-to-height ratio	6324	<0.001**	continuous normal	ALL
7	Hypertension	6324	0.249	categorical	ALL
8	Type-2 diabetes	6324	0.017**	categorical	ALL
9	Dyslipidemia	6324	0.423	categorical	ALL
10	Family history of premature CHD	6324	0.581	categorical	ALL
11	Hormone-replacement therapy	5661	0.850	categorical	ALL
12	MeDiet Adherence score	6324	<0.001**	continuous normal	ALL
13	follow-up to main event (years)	6324	<0.001**	continuous normal	ALL
14	AMI, stroke, or CV Death	6324	0.064*	categorical	ALL

-----

Signif. codes: 0 '\*\*' 0.05 '\*' 0.1 ' ' 1

- If you are interested in describing all variables but a subset of them use '-' (this is useful when having variables such as 'id', 'hc', 'name', ...)

```
descr2 <- compareGroups(group ~ . -sex -age -event, predimed)
descr2
```

----- Summary of results by groups of 'Intervention group'-----

	var	N	p.value	method	selection
1	Smoking	6324	0.444	categorical	ALL
2	Body mass index	6324	<0.001**	continuous normal	ALL
3	Waist circumference	6324	0.045**	continuous normal	ALL
4	Waist-to-height ratio	6324	<0.001**	continuous normal	ALL
5	Hypertension	6324	0.249	categorical	ALL
6	Type-2 diabetes	6324	0.017**	categorical	ALL
7	Dyslipidemia	6324	0.423	categorical	ALL
8	Family history of premature CHD	6324	0.581	categorical	ALL
9	Hormone-replacement therapy	5661	0.850	categorical	ALL
10	MeDiet Adherence score	6324	<0.001**	continuous normal	ALL
11	follow-up to main event (years)	6324	<0.001**	continuous normal	ALL

-----

Signif. codes: 0 '\*\*\*' 0.05 '\*' 0.1 ' ' 1

### 3. Build the descriptive table.

```
descrtable <- createTable(descr)
descrtable
```

-----Summary descriptives table by 'Intervention group'-----

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + VOO N=2182	p.overall
Sex:				<0.001
Male	812 (39.8%)	968 (46.1%)	899 (41.2%)	
Female	1230 (60.2%)	1132 (53.9%)	1283 (58.8%)	
Age	67.3 (6.28)	66.7 (6.02)	67.0 (6.21)	0.003
Smoking:				0.444
Never	1282 (62.8%)	1259 (60.0%)	1351 (61.9%)	
Current	270 (13.2%)	296 (14.1%)	292 (13.4%)	
Former	490 (24.0%)	545 (26.0%)	539 (24.7%)	
Body mass index	30.3 (3.96)	29.7 (3.77)	29.9 (3.71)	<0.001
Waist circumference	101 (10.8)	100 (10.6)	100 (10.4)	0.045
Waist-to-height ratio	0.63 (0.07)	0.62 (0.06)	0.63 (0.06)	<0.001
Hypertension:				0.249
No	331 (16.2%)	362 (17.2%)	396 (18.1%)	
Yes	1711 (83.8%)	1738 (82.8%)	1786 (81.9%)	
Type-2 diabetes:				0.017
No	1072 (52.5%)	1150 (54.8%)	1100 (50.4%)	
Yes	970 (47.5%)	950 (45.2%)	1082 (49.6%)	
Dyslipidemia:				0.423
No	563 (27.6%)	561 (26.7%)	622 (28.5%)	
Yes	1479 (72.4%)	1539 (73.3%)	1560 (71.5%)	
Family history of premature CHD:				0.581

# Customizing results

- Hide 'No' category

```
update(descrtable, hide.no='no')
```

-----Summary descriptives table by 'Intervention group'-----

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + V00 N=2182	p.overall
Sex:				<0.001
Male	812 (39.8%)	968 (46.1%)	899 (41.2%)	
Female	1230 (60.2%)	1132 (53.9%)	1283 (58.8%)	
Age	67.3 (6.28)	66.7 (6.02)	67.0 (6.21)	0.003
Smoking:				0.444
Never	1282 (62.8%)	1259 (60.0%)	1351 (61.9%)	
Current	270 (13.2%)	296 (14.1%)	292 (13.4%)	
Former	490 (24.0%)	545 (26.0%)	539 (24.7%)	
Body mass index	30.3 (3.96)	29.7 (3.77)	29.9 (3.71)	<0.001
Waist circumference	101 (10.8)	100 (10.6)	100 (10.4)	0.045
Waist-to-height ratio	0.63 (0.07)	0.62 (0.06)	0.63 (0.06)	<0.001
Hypertension	1711 (83.8%)	1738 (82.8%)	1786 (81.9%)	0.249
Type-2 diabetes	970 (47.5%)	950 (45.2%)	1082 (49.6%)	0.017
Dyslipidemia	1479 (72.4%)	1539 (73.3%)	1560 (71.5%)	0.423
Family history of premature CHD	462 (22.6%)	460 (21.9%)	507 (23.2%)	0.581
Hormone-replacement therapy	31 (1.68%)	30 (1.61%)	36 (1.84%)	0.850
MeDiet Adherence score	8.44 (1.94)	8.81 (1.90)	8.77 (1.97)	<0.001
follow-up to main event (years)	4.09 (1.74)	4.31 (1.70)	4.64 (1.60)	<0.001
AMI, stroke, or CV Death	97 (4.75%)	70 (3.33%)	85 (3.90%)	0.064

► Show number of valid data

```
update(descrtable, hide.no='no', show.n = TRUE)
```

-----Summary descriptives table by 'Intervention group'-----

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + VOO N=2182	p.overall	N
Sex:				<0.001	6324
Male	812 (39.8%)	968 (46.1%)	899 (41.2%)		
Female	1230 (60.2%)	1132 (53.9%)	1283 (58.8%)		
Age	67.3 (6.28)	66.7 (6.02)	67.0 (6.21)	0.003	6324
Smoking:				0.444	6324
Never	1282 (62.8%)	1259 (60.0%)	1351 (61.9%)		
Current	270 (13.2%)	296 (14.1%)	292 (13.4%)		
Former	490 (24.0%)	545 (26.0%)	539 (24.7%)		
Body mass index	30.3 (3.96)	29.7 (3.77)	29.9 (3.71)	<0.001	6324
Waist circumference	101 (10.8)	100 (10.6)	100 (10.4)	0.045	6324
Waist-to-height ratio	0.63 (0.07)	0.62 (0.06)	0.63 (0.06)	<0.001	6324
Hypertension	1711 (83.8%)	1738 (82.8%)	1786 (81.9%)	0.249	6324
Type-2 diabetes	970 (47.5%)	950 (45.2%)	1082 (49.6%)	0.017	6324
Dyslipidemia	1479 (72.4%)	1539 (73.3%)	1560 (71.5%)	0.423	6324
Family history of premature CHD	462 (22.6%)	460 (21.9%)	507 (23.2%)	0.581	6324
Hormone-replacement therapy	31 (1.68%)	30 (1.61%)	36 (1.84%)	0.850	5661
MedDiet Adherence score	8.44 (1.94)	8.81 (1.90)	8.77 (1.97)	<0.001	6324
follow-up to main event (years)	4.09 (1.74)	4.31 (1.70)	4.64 (1.60)	<0.001	6324
AMI, stroke, or CV Death	97 (4.75%)	70 (3.33%)	85 (3.90%)	0.064	6324

► Show only relative percentages

```
update(descrtable, hide.no='no', show.n = TRUE, type=1)
```

-----Summary descriptives table by 'Intervention group'-----

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + VOO N=2182	p.overall	N
Sex:				<0.001	6324
Male	39.8%	46.1%	41.2%		
Female	60.2%	53.9%	58.8%		
Age	67.3 (6.28)	66.7 (6.02)	67.0 (6.21)	0.003	6324
Smoking:				0.444	6324
Never	62.8%	60.0%	61.9%		
Current	13.2%	14.1%	13.4%		
Former	24.0%	26.0%	24.7%		
Body mass index	30.3 (3.96)	29.7 (3.77)	29.9 (3.71)	<0.001	6324
Waist circumference	101 (10.8)	100 (10.6)	100 (10.4)	0.045	6324
Waist-to-height ratio	0.63 (0.07)	0.62 (0.06)	0.63 (0.06)	<0.001	6324
Hypertension	83.8%	82.8%	81.9%	0.249	6324
Type-2 diabetes	47.5%	45.2%	49.6%	0.017	6324
Dyslipidemia	72.4%	73.3%	71.5%	0.423	6324
Family history of premature CHD	22.6%	21.9%	23.2%	0.581	6324
Hormone-replacement therapy	1.68%	1.61%	1.84%	0.850	5661
MeDiet Adherence score	8.44 (1.94)	8.81 (1.90)	8.77 (1.97)	<0.001	6324
follow-up to main event (years)	4.09 (1.74)	4.31 (1.70)	4.64 (1.60)	<0.001	6324
AMI, stroke, or CV Death	4.75%	3.33%	3.90%	0.064	6324

## Customizing descriptives (tests)

- By default, compareGroups report means and SD, and performs t-test or ANOVA for continuous variables.
- To report medians and quartiles and perform Kruskal-Wallis tests for continuous variable:

```
descr <- update(descr, method=2)
createTable(descr, hide.no="no")
```

-----Summary descriptives table by 'Intervention group'-----

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + V00 N=2182	p.overall
Sex:				<0.001
Male	812 (39.8%)	968 (46.1%)	899 (41.2%)	
Female	1230 (60.2%)	1132 (53.9%)	1283 (58.8%)	
Age	67.0 [62.0;72.0]	66.0 [62.0;71.0]	67.0 [62.0;72.0]	0.003
Smoking:				0.444
Never	1282 (62.8%)	1259 (60.0%)	1351 (61.9%)	
Current	270 (13.2%)	296 (14.1%)	292 (13.4%)	
Former	490 (24.0%)	545 (26.0%)	539 (24.7%)	
Body mass index	30.0 [27.5;32.8]	29.5 [26.9;32.2]	29.7 [27.2;32.4]	<0.001
Waist circumference	101 [94.0;108]	100 [93.0;107]	100 [93.0;107]	0.085
Waist-to-height ratio	0.63 [0.59;0.68]	0.62 [0.58;0.66]	0.62 [0.58;0.67]	<0.001
Hypertension	1711 (83.8%)	1738 (82.8%)	1786 (81.9%)	0.249

► Change number of decimals

```
update(descrtable, hide.no='no', digits=1, digits.p=5)
```

-----Summary descriptives table by 'Intervention group'-----

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + V00 N=2182	p.overall
Sex:				0.00008
Male	812 (39.8%)	968 (46.1%)	899 (41.2%)	
Female	1230 (60.2%)	1132 (53.9%)	1283 (58.8%)	
Age	67.0 [62.0;72.0]	66.0 [62.0;71.0]	67.0 [62.0;72.0]	0.00299
Smoking:				0.44435
Never	1282 (62.8%)	1259 (60.0%)	1351 (61.9%)	
Current	270 (13.2%)	296 (14.1%)	292 (13.4%)	
Former	490 (24.0%)	545 (26.0%)	539 (24.7%)	
Body mass index	30.0 [27.5;32.8]	29.5 [26.9;32.2]	29.7 [27.2;32.4]	0.00002
Waist circumference	101.0 [94.0;108.0]	100.0 [93.0;107.0]	100.0 [93.0;107.0]	0.08460
Waist-to-height ratio	0.6 [0.6;0.7]	0.6 [0.6;0.7]	0.6 [0.6;0.7]	0.00019
Hypertension	1711 (83.8%)	1738 (82.8%)	1786 (81.9%)	0.24876
Type-2 diabetes	970 (47.5%)	950 (45.2%)	1082 (49.6%)	0.01725
Dyslipidemia	1479 (72.4%)	1539 (73.3%)	1560 (71.5%)	0.42297
Family history of premature CHD	462 (22.6%)	460 (21.9%)	507 (23.2%)	0.58131
Hormone-replacement therapy	31 (1.7%)	30 (1.6%)	36 (1.8%)	0.85009
MeDiet Adherence score	8.0 [7.0;10.0]	9.0 [8.0;10.0]	9.0 [8.0;10.0]	<0.00001
follow-up to main event (years)	4.2 [2.7;5.6]	4.7 [2.8;5.8]	5.0 [3.4;5.9]	<0.00001
AMI, stroke, or CV Death	97 (4.8%)	70 (3.3%)	85 (3.9%)	0.06386



► Perform medians and quantiles for some variables:

```
descr <- update(descr, method=c(age=2, p14=2))
createTable(descr, hide.no="no")
```

-----Summary descriptives table by 'Intervention group'-----

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + V00 N=2182	p.overall
Sex:				<0.001
Male	812 (39.8%)	968 (46.1%)	899 (41.2%)	
Female	1230 (60.2%)	1132 (53.9%)	1283 (58.8%)	
Age	67.0 [62.0;72.0]	66.0 [62.0;71.0]	67.0 [62.0;72.0]	0.003
Smoking:				0.444
Never	1282 (62.8%)	1259 (60.0%)	1351 (61.9%)	
Current	270 (13.2%)	296 (14.1%)	292 (13.4%)	
Former	490 (24.0%)	545 (26.0%)	539 (24.7%)	
Body mass index	30.3 (3.96)	29.7 (3.77)	29.9 (3.71)	<0.001
Waist circumference	101 (10.8)	100 (10.6)	100 (10.4)	0.045
Waist-to-height ratio	0.63 (0.07)	0.62 (0.06)	0.63 (0.06)	<0.001
Hypertension	1711 (83.8%)	1738 (82.8%)	1786 (81.9%)	0.249
Type-2 diabetes	970 (47.5%)	950 (45.2%)	1082 (49.6%)	0.017
Dyslipidemia	1479 (72.4%)	1539 (73.3%)	1560 (71.5%)	0.423
Family history of premature CHD	462 (22.6%)	460 (21.9%)	507 (23.2%)	0.581
Hormone-replacement therapy	31 (1.68%)	30 (1.61%)	36 (1.84%)	0.850
MedDiet Adherence score	8.00 [7.00;10.0]	9.00 [8.00;10.0]	9.00 [8.00;10.0]	<0.001
follow-up to main event (years)	4.09 (1.74)	4.31 (1.70)	4.64 (1.60)	<0.001
AMI, stroke, or CV Death	97 (4.75%)	70 (3.33%)	85 (3.90%)	0.064

# Odds Ratio

- ▶ Place the case/control variable on left hand side.
- ▶ It computes the Odds Ratio (OR) of being a case (second category). To change reference category, use `ref.y` argument from `compareGroups` function.
- ▶ Let's report the OR of being hyperchol

```
table(predimed$hyperchol)
```

No	Yes
1746	4578

```
descr <- compareGroups(hyperchol ~ ., predimed)
createTable(descr, hide.no="no", show.ratio=TRUE,
            show.p.overall=FALSE, show.p.trend = FALSE)
```

-----Summary descriptives table by 'Dyslipidemia'-----

	No N=1746	Yes N=4578	OR	p.ratio
Intervention group:				
Control	563 (32.2%)	1479 (32.3%)	Ref.	Ref.
MedDiet + Nuts	561 (32.1%)	1539 (33.6%)	1.04 [0.91;1.20]	0.536
MedDiet + VOO	622 (35.6%)	1560 (34.1%)	0.95 [0.83;1.09]	0.499
Sex:				
Male	906 (51.9%)	1773 (38.7%)	Ref.	Ref.
Female	840 (48.1%)	2805 (61.3%)	1.71 [1.53;1.91]	0.000
Age	67.6 (6.23)	66.8 (6.14)	0.98 [0.97;0.99]	<0.001
Smoking:				
Never	980 (56.1%)	2912 (63.6%)	Ref.	Ref.
Current	291 (16.7%)	567 (12.4%)	0.66 [0.56;0.77]	<0.001
Former	475 (27.2%)	1099 (24.0%)	0.78 [0.68;0.89]	<0.001
Body mass index	30.0 (3.89)	29.9 (3.79)	0.99 [0.98;1.01]	0.353
Waist circumference	101 (10.4)	100.0 (10.6)	0.99 [0.98;0.99]	<0.001
Waist-to-height ratio	0.63 (0.07)	0.63 (0.07)	0.49 [0.21;1.13]	0.096
Hypertension	1337 (76.6%)	3898 (85.1%)	1.75 [1.53;2.01]	<0.001
Type-2 diabetes	1222 (70.0%)	1780 (38.9%)	0.27 [0.24;0.31]	0.000
Family history of premature CHD	409 (23.4%)	1020 (22.3%)	0.94 [0.82;1.07]	0.331
Hormone-replacement therapy	26 (1.65%)	71 (1.74%)	1.05 [0.67;1.68]	0.844
MeDiet Adherence score	8.68 (1.90)	8.68 (1.96)	1.00 [0.97;1.03]	0.995
follow-up to main event (years)	4.59 (1.63)	4.26 (1.71)	0.89 [0.86;0.92]	<0.001
AMI, stroke, or CV Death	101 (5.78%)	151 (3.30%)	0.56 [0.43;0.72]	<0.001

# Hazard Ratios

- ▶ PREDIMED is a cohort study with time-to-event outcome.
- ▶ Descriptives by cases and controls, HR taking into account time-to-event response (with possible right censoring) and p-values are easily computed

## 1. First, create a Surv variable

```
predimed$tevent <- with(predimed, Surv(toevent, event=="Yes"))
```

## 2. Then write this variable on left side of ~ in compareGroups.

Note the use of - to erase some variables.

```
descr <- compareGroups(tevent ~ .- toevent-event, predimed)
createTable(descr, hide.no="no", show.ratio=TRUE,
             show.p.overall=FALSE)
```

-----Summary descriptives table by 'tevent'-----

	No event N=6072	Event N=252	HR	p.ratio
Intervention group:				
Control	1945 (32.0%)	97 (38.5%)	Ref.	Ref.
MedDiet + Nuts	2030 (33.4%)	70 (27.8%)	0.66 [0.48;0.89]	0.008
MedDiet + VOO	2097 (34.5%)	85 (33.7%)	0.70 [0.53;0.94]	0.018
Sex:				
Male	2528 (41.6%)	151 (59.9%)	Ref.	Ref.
Female	3544 (58.4%)	101 (40.1%)	0.49 [0.38;0.63]	<0.001
Age	66.9 (6.14)	69.4 (6.65)	1.06 [1.04;1.09]	<0.001
Smoking:				
Never	3778 (62.2%)	114 (45.2%)	Ref.	Ref.
Current	809 (13.3%)	49 (19.4%)	1.96 [1.40;2.74]	<0.001

# Utilities

- use label function from Hmisc package to label variables

```
Hmisc::label(predimed$age) <- "Age of participant"
```

- To know the original variable names (instead of labels)

```
descrtable <- createTable(compareGroups(group ~ ., predimed))  
varinfo(descrtable)
```

--- Analyzed variable names ---

	Orig varname	Shown varname
1	group	Intervention group
2	sex	Sex
3	age	Age of participant
4	smoke	Smoking
5	bmi	Body mass index
6	waist	Waist circumference
7	wth	Waist-to-height ratio
8	htn	Hypertension
9	diab	Type-2 diabetes
10	hyperchol	Dyslipidemia
11	famhist	Family history of premature CHD
12	hormo	Hormone-replacement therapy
13	p14	MeDiet Adherence score
14	toevent	follow-up to main event (years)

- To select some variables use [], indexing by names or by position

```
descrtable <- createTable(compareGroups(group ~ ., predimed))
descrtable[c('age', 'bmi')]
```

-----Summary descriptives table by 'Intervention group'-----

	Control N=2042	MedDiet + Nuts N=2100	MedDiet + V00 N=2182	p.overall
Age of participant	67.3 (6.28)	66.7 (6.02)	67.0 (6.21)	0.003
Body mass index	30.3 (3.96)	29.7 (3.77)	29.9 (3.71)	<0.001

```
descrtable[c(1,4)]
```

-----Summary descriptives table by 'Intervention group'-----

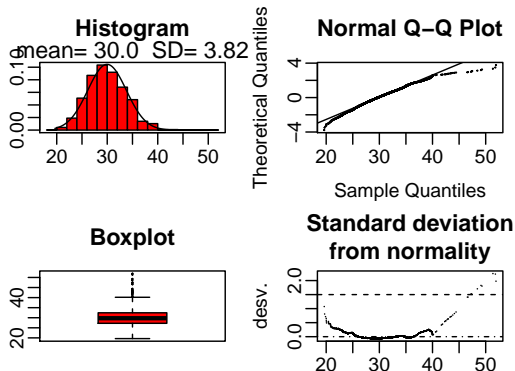
	Control N=2042	MedDiet + Nuts N=2100	MedDiet + V00 N=2182	p.overall
Sex:				<0.001
Male	812 (39.8%)	968 (46.1%)	899 (41.2%)	
Female	1230 (60.2%)	1132 (53.9%)	1283 (58.8%)	
Body mass index	30.3 (3.96)	29.7 (3.77)	29.9 (3.71)	<0.001

# Plotting variables

## ► Continuous univariate

```
descr <- compareGroups(group ~ ., predimed)
plot(descr['bmi'])
```

## Normality plots of 'Body mass index'

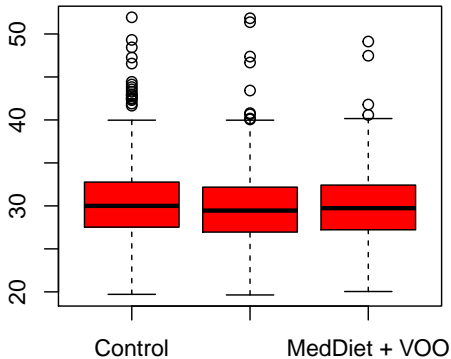




► Continuous by groups

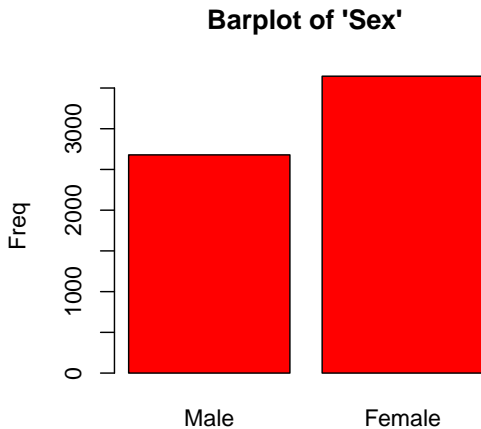
```
plot(descr['bmi'], bivar=TRUE)
```

boxplot of 'Body mass index' by 'Intervention'



- Categorical variable

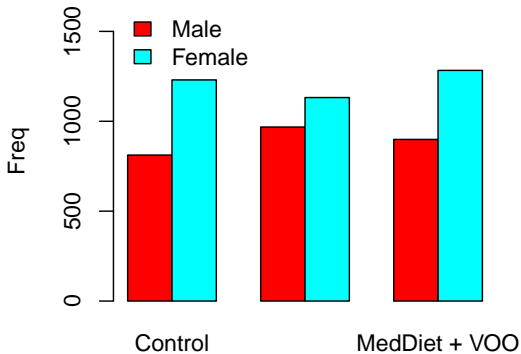
```
plot(descr['sex'])
```



► Categorical by groups

```
plot(descr['sex'], bivar=TRUE)
```

**Barplot of 'Sex' by 'Intervention group'**



# Export

```
# CSV
export2csv(descrtable, file="tabla.csv", sep=";")
# Excel
export2xls(descrtable, file="tabla.xlsx")
# Word
export2word(descrtable, file="tabla.docx")
# Latex
export2tex(descrtable, file="tabla.tex")
```

... or inside a **Rmarkdown** document chunk

```
export2md(descrtable)
```

Table 1: Summary descriptives table by groups o

	Control N=2042	MedDiet + Nuts
Sex:		
Male	812 (39.8%)	968 (46.1%)
Female	1230 (60.2%)	1132 (53.9%)
Age of participant	67.3 (6.28)	66.7 (6.02)

## More

- ▶ There exists much more options
- ▶ See `?compareGroups`, `?createTable`, ...
- ▶ Visit `compareGroups` webpage
- ▶ Application made with Shiny available here

Step 1. Load data

Step 2. Select variables

Selected

group  
sex  
age  
smoke  
bmi  
waist  
wth  
htn  
diab  
hyperchol  
famhist  
hormo  
p14  
toevent  
event

Discarded

Update

Step 3. Settings

Step 4. Display

Step 5. Save table

ABOUT DATA - TABLE PLOT SNPs

View options (Show)

	[ALL] N=6324	Control N=2042	MedDiet + Nuts N=2100	MedDiet + VOO N=2182	p.overall
Intervention group:					0.000
Control	2042 (32.3%)	2042 (100%)	0 (0.00%)	0 (0.00%)	
MedDiet + Nuts	2100 (33.2%)	0 (0.00%)	2100 (100%)	0 (0.00%)	
MedDiet + VOO	2182 (34.5%)	0 (0.00%)	0 (0.00%)	2182 (100%)	
Sex:					<0.001
Male	2679 (42.4%)	812 (39.8%)	968 (46.1%)	899 (41.2%)	
Female	3645 (57.6%)	1230 (60.2%)	1132 (53.9%)	1283 (58.8%)	
Age	67.0 [62.0;72.0]	67.0 [62.0;72.0]	66.0 [62.0;71.0]	67.0 [62.0;72.0]	0.003
Smoking:					0.444
Never	3892 (61.5%)	1282 (62.8%)	1259 (60.0%)	1351 (61.9%)	
Current	858 (13.6%)	270 (13.2%)	296 (14.1%)	292 (13.4%)	
Former	1574 (24.9%)	490 (24.0%)	545 (26.0%)	539 (24.7%)	
Body mass index	29.8 [27.2;32.5]	30.0 [27.5;32.8]	29.5 [26.9;32.2]	29.7 [27.2;32.4]	<0.001
Waist circumference	100 [93.0;107]	101 [94.0;108]	100 [93.0;107]	100 [93.0;107]	0.085
Waist-to-height ratio	0.63 [0.58;0.67]	0.63 [0.59;0.68]	0.62 [0.58;0.66]	0.62 [0.58;0.67]	<0.001
Hypertension	5235 (82.8%)	1711 (83.8%)	1738 (82.8%)	1786 (81.9%)	0.249
Type-2 diabetes	3002 (47.5%)	970 (47.5%)	950 (45.2%)	1082 (49.6%)	0.017
Dyslipidemia	4578 (72.4%)	1479 (72.4%)	1539 (73.3%)	1560 (71.5%)	0.423
Family history of premature CHD	1429 (22.6%)	462 (22.6%)	460 (21.9%)	507 (23.2%)	0.581
Hormone-replacement therapy	97 (1.71%)	31 (1.68%)	30 (1.61%)	36 (1.84%)	0.850

Figure 3: compareGroups Shiny app

30 / 31

# Session info

## sessionInfo()

R version 3.5.0 (2018-04-23)

Platform: x86\_64-w64-mingw32/x64 (64-bit)

Running under: Windows 10 x64 (build 17134)

Matrix products: default

locale:

[1] LC\_COLLATE=Spanish\_Spain.1252 LC\_CTYPE=Spanish\_Spain.1252

[3] LC\_MONETARY=Spanish\_Spain.1252 LC\_NUMERIC=C

[5] LC\_TIME=Spanish\_Spain.1252

attached base packages:

[1] stats graphics grDevices utils datasets methods base

other attached packages:

[1] compareGroups\_3.4.0 SNPassoc\_1.9-8 mvtnorm\_1.0-7

[4] haplo.stats\_1.7.9 xtable\_1.8-2 gdata\_2.18.0

[7] Hmisc\_4.1-1 ggplot2\_3.0.0 Formula\_1.2-3

[10] survival\_2.41-3 lattice\_0.20-35 knitr\_1.20

loaded via a namespace (and not attached):

[1] nlme\_3.1-137 bitops\_1.0-6 rms\_5.1-2

[4] bit64\_0.9-7 RColorBrewer\_1.1-2 progress\_1.1.2

[7] httr\_1.3.1 rprojroot\_1.3-2 tools\_3.5.0

[10] backports\_1.1.2 R6\_2.2.2 rpart\_4.1-13

[13] DBI\_1.0.0 lazyeval\_0.2.1 BiocGenerics\_0.26.0

[16] colorspace\_1.3-2 jomo\_2.6-2 nnet\_7.3-12

[19] withr\_2.1.2 mnormt\_1.5-5 tidyselect\_0.2.4

[22] gridExtra\_2.3 prettyunits\_1.0.2 bit\_1.1-13