Trainer: Bondor Cosmina-Ioana, PhD, MsC

Descriptive statistics

Qualitative Variables

A ALWAYS

S SEEK

K KNOWLEDGE

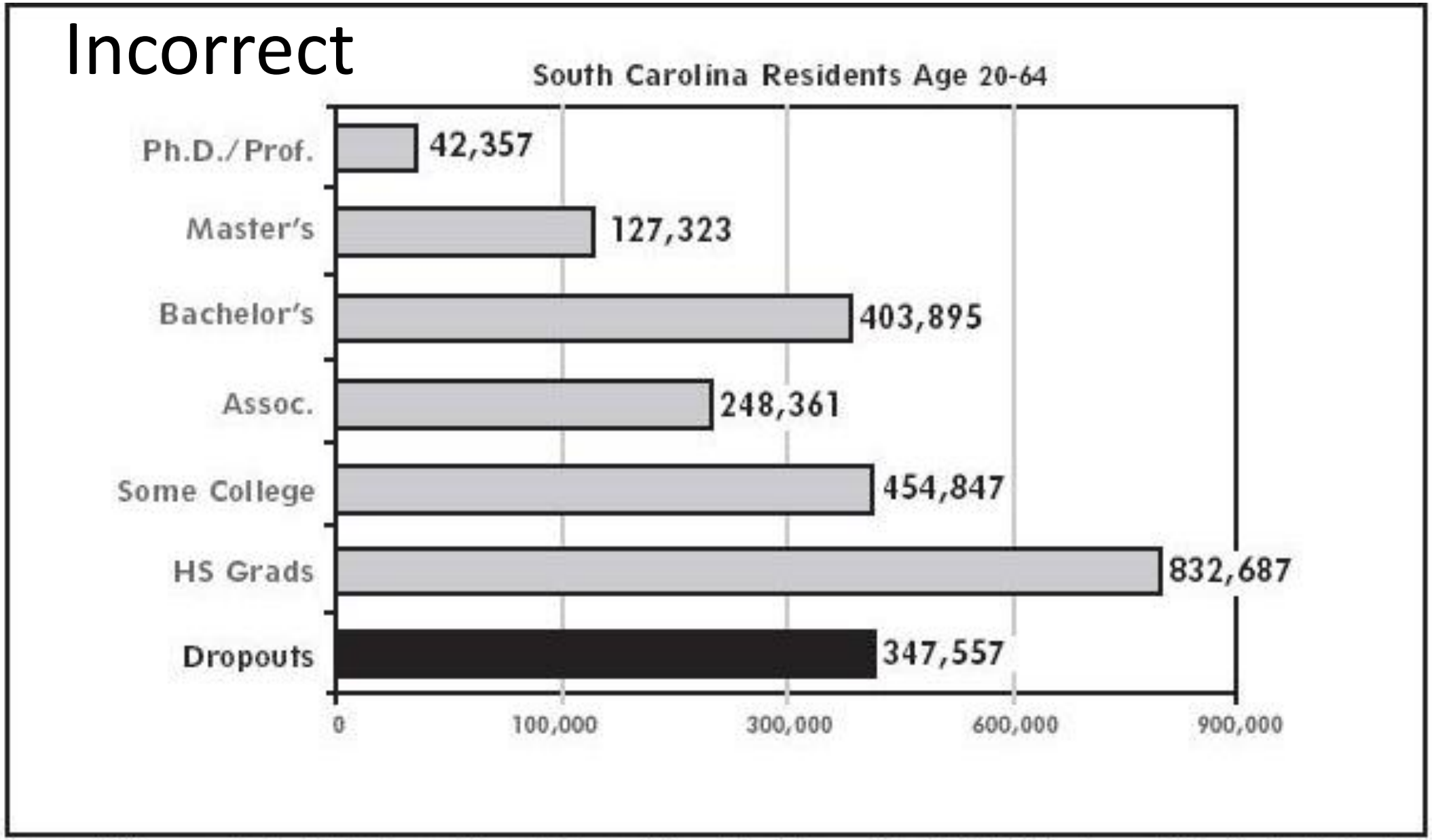
Objectives for descriptive statistics

- Scales of measurement
- Variables type
- Frequencies
- Indicators for qualitative variables
- Relationships between two variables

} Descriptive statistics 2

- Dispersion
- Ways to asses the reliability and the validity of measurements
- Misleading charts examples
- Exercises

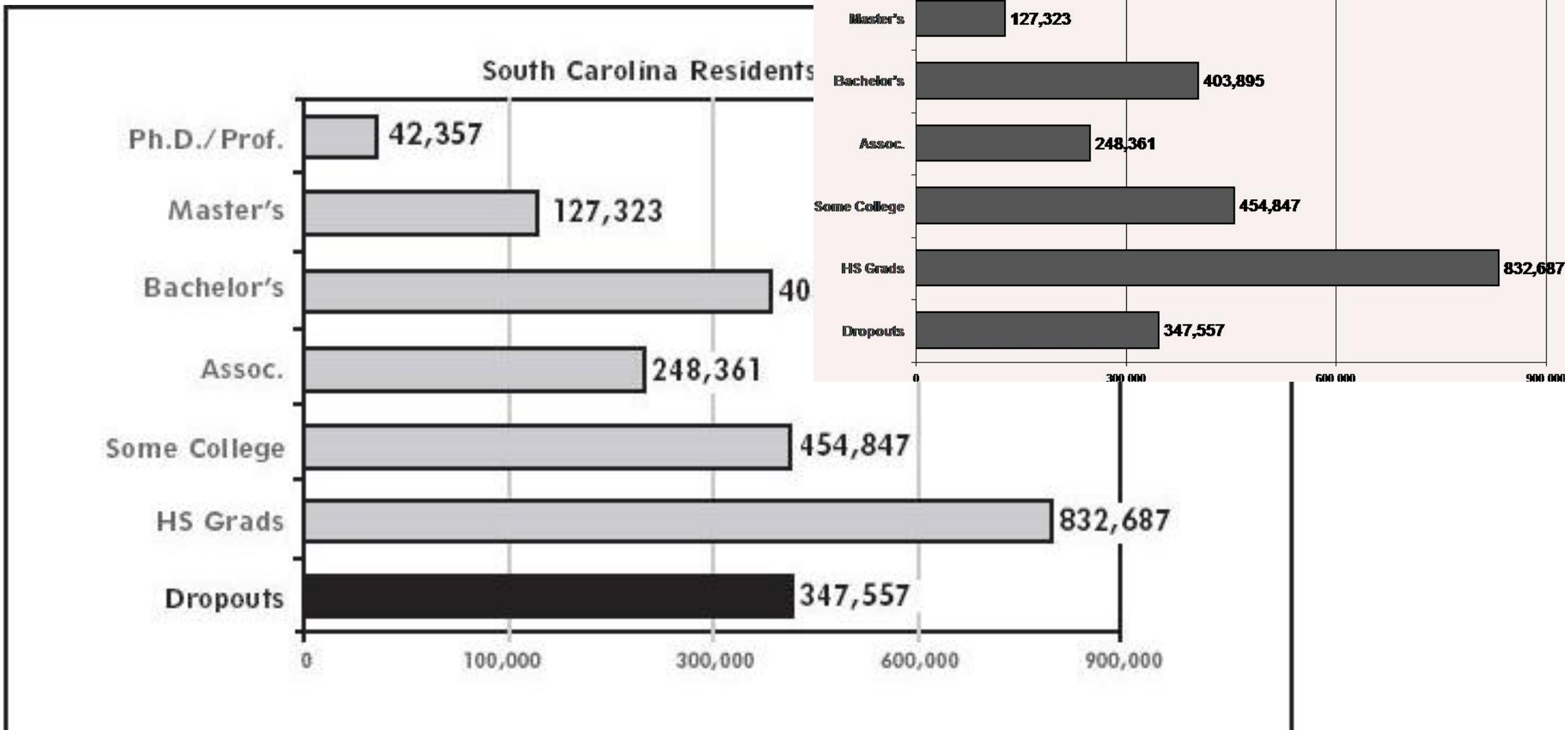
Incorrect

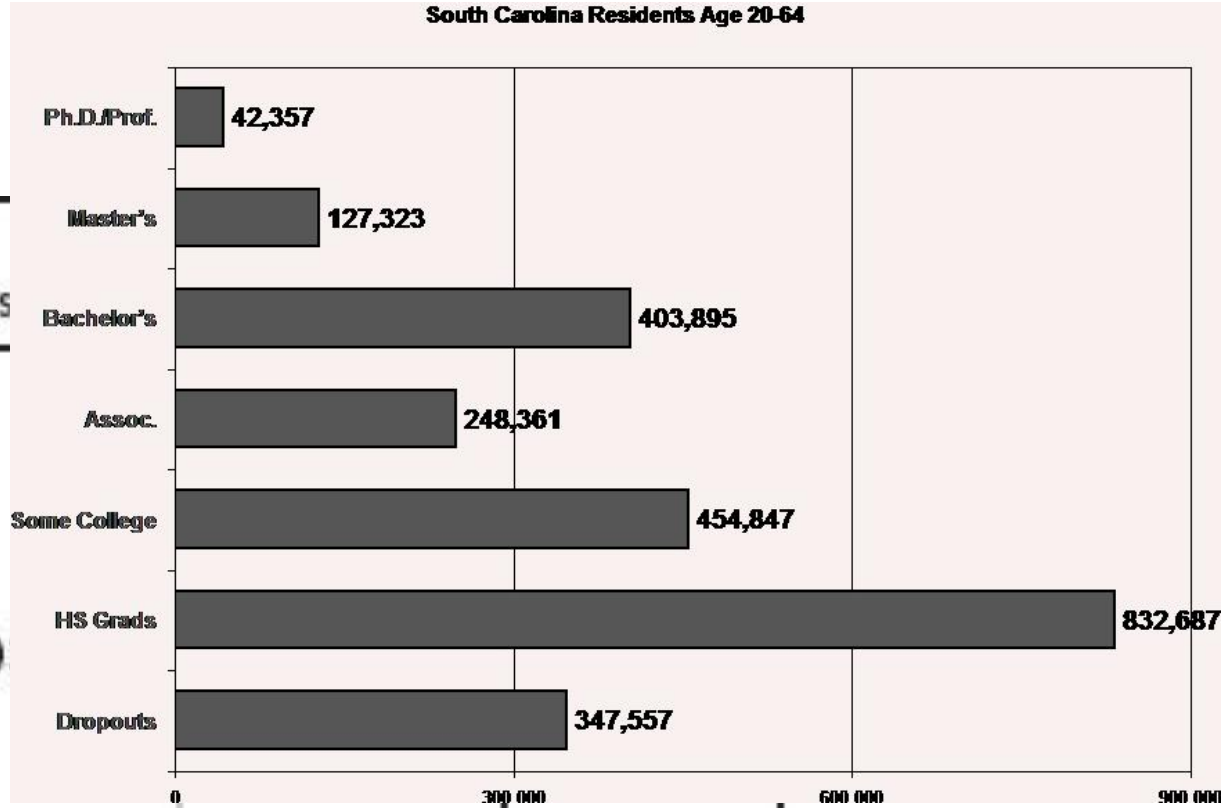


Source: PolEcon analysis of U.S. Census Bureau's Current Population Survey March 2004-06 Supplement data for South Carolina

http://www.shermandorn.com/mt/archives/001150.html

Incorrect Ox axis





Two Qualitative Variables

How we understand and organize the data when we investigate two qualitative variables?

Contingency Tables

– the way data are organized when we have two qualitative variables in some relationship

Why Do We Need Contingency Tables?

- 2 variables: stress and sleeping (≥8hours)

- Questions:
  - "Are you feeling stressed?" Yes/No
  - "Did you sleep enough last two-three nights (≥8hours)?" Yes/No

- Survey 100 subjects

- Ex. Is Air Polution (Yes/No) a risk factor for Asthma (Present /Absent)
- Is the treatment with rituximab (Yes/No) efficient for monoclonal gammopathy (Present/Absent)?
- Is the ultrasonography (Yes/No) a diagnostic procedure for Pregnancy (Yes/No)

- 2 variables: stress and sleeping (≥8hours)
- Questions:
  - “Are you feeling stressed?“ Yes/No
  - “Did you sleep enough last two-three nights (≥8hours)?”
- Survey 100 subjects

Id_patient	Stress	Sleeping (≥8hours)
1	Yes	Yes
2	Yes	Yes
3	Yes	Yes
4	Yes	Yes
5	Yes	No
6	Yes	No
7	Yes	No
8	Yes	No
9	Yes	No
10	Yes	No
11	No	Yes
12	No	Yes
13	No	Yes
14	No	Yes
15	No	Yes
16	No	Yes
17	No	Yes
18	No	Yes
19	No	Yes
20	No	Yes

What Is a Contingency Table?

- a frequency table to organize data for two qualitative variables when we want to show the joint

Shows:

- **Joint frequencies**
- **Totals on columns/rows**
- **Grand total**

What Is a Contingency Table?

- a frequency table to organize data for two qualitative variables

Shows:

- **Joint frequencies**
- **Totals on columns/rows**
- **Grand total**

	Variable 2 – category 1	Varable 2 – category 2	Total
Variable 1 – category 1	a	b	a+b
Variable 1 – category 2	c	d	c+d
Total	a+c	b+d	n=a+b+c+d

Contingency table for two dichotomial (binary) variable

What Is a Contingency Table?

- a frequency table to organize data for two qualitative variables

Shows:

- **Joint frequencies**
- **Totals on columns/rows**
- **Grand total**

	Variable 2 – category 1	Varable 2 – category 2	Total
Variable 1 – category 1	a	b	a+b
Variable 1 – category 2	c	d	c+d
Total	a+c	b+d	n=a+b+c+d

Contingency table for two dichotomial (binary) variable

Two Dichotomial Variables

Id_patient	Stress	Sleeping (≥8hours)
1	Yes	Yes
2	Yes	Yes
3	Yes	Yes
4	Yes	Yes
5	Yes	No
6	Yes	No
7	Yes	No
8	Yes	No
9	Yes	No
10	Yes	No
11	No	Yes
12	No	Yes
13	No	Yes
14	No	Yes
15	No	Yes
16	No	Yes
17	No	Yes

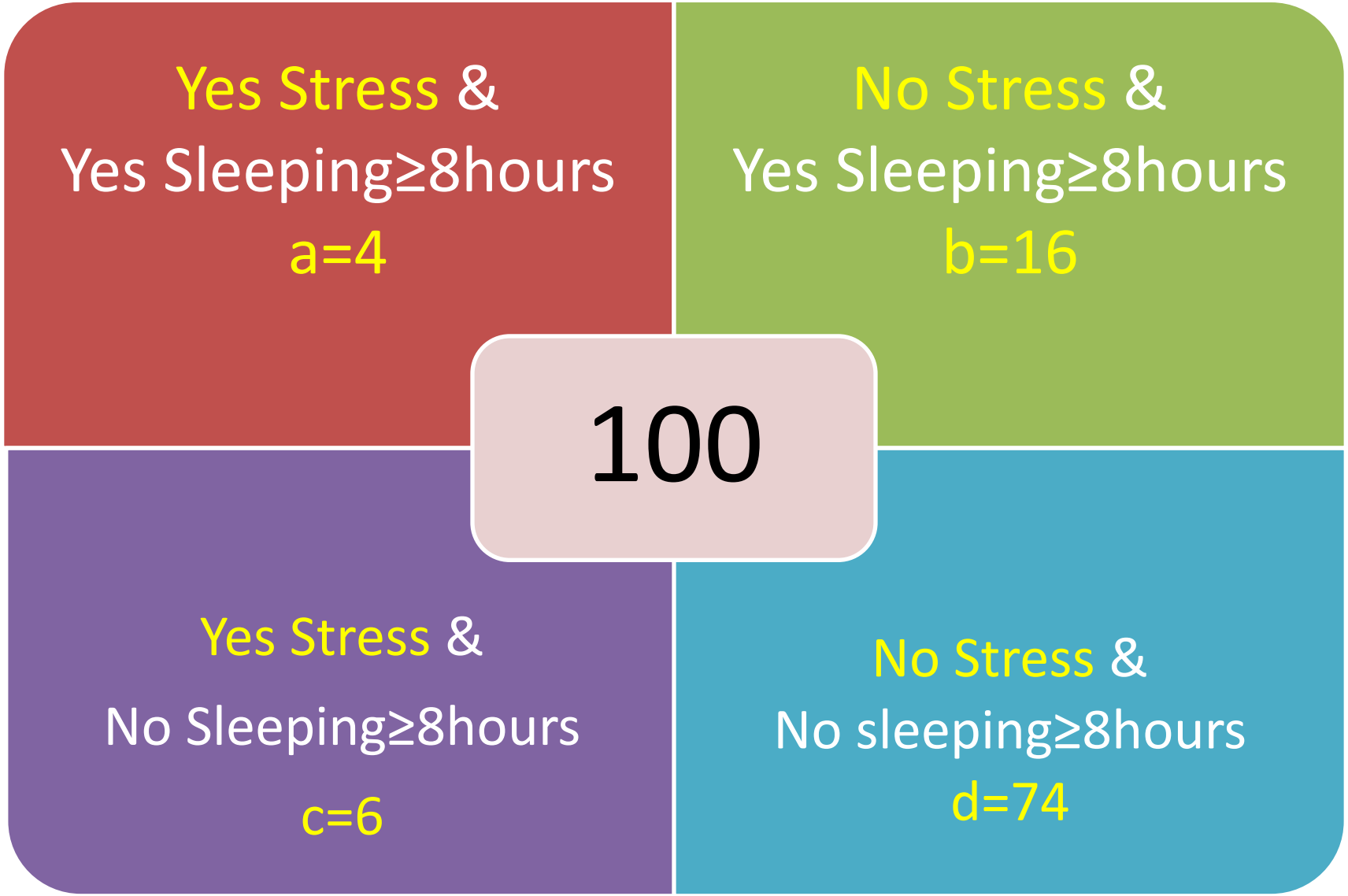
"Are you feeling stressed?"

"Did you sleep enough last two-three nights (≥8hours)?"

	Stress **Yes**	Stress **No**	Total
Sleeping (≥8hours) **Yes**	4	16	20
Sleeping (≥8hours) **No**	6	74	80
Total	10	90	100

Contingency table for two dichotomial (binary) variable

2 dichotomial variables both with 2 categories → 4 possible combined categories



Two Dichotomial Variables

"Are you feeling stressed?"

Stress	Frequency
Yes	10
No	90
Total	100

"Did you sleep enough last two-three nights (≥8hours)?"

Sleeping (≥8hours)	Frequency
Yes	20
No	80
Total	100

	Stress **Yes**	Stress **No**	Total
Sleeping (≥8hours) **Yes**	4	16	20
Sleeping (≥8hours) **No**	6	74	80
Total	10	90	100

Contingency table for two dichotomial (binary) variable

Two Dichotomial Variables

	Stress Yes	Stress No	Total
Sleeping (≥8hours) Yes	4	16	20
Sleeping (≥8hours) No	6	74	80
Total	10	90	100

Stress	Frequency
Yes	10
No	90
Total	100

Sleeping (≥8hours)	Frequency
Yes	20
No	80
Total	100

Contingency table – frequency table for presenting the association of two dichotomial variables

	Stress **Yes**	Stress **No**	Total
Sleeping (≥8hours) **Yes**	4	16	20
Sleeping (≥8hours) **No**	6	74	80
Total	10	90	100

Joint absolute frequency =

People who respond that they are stressed & they sleep more then 8 hours = 4

Total on column =10

Total on row =20

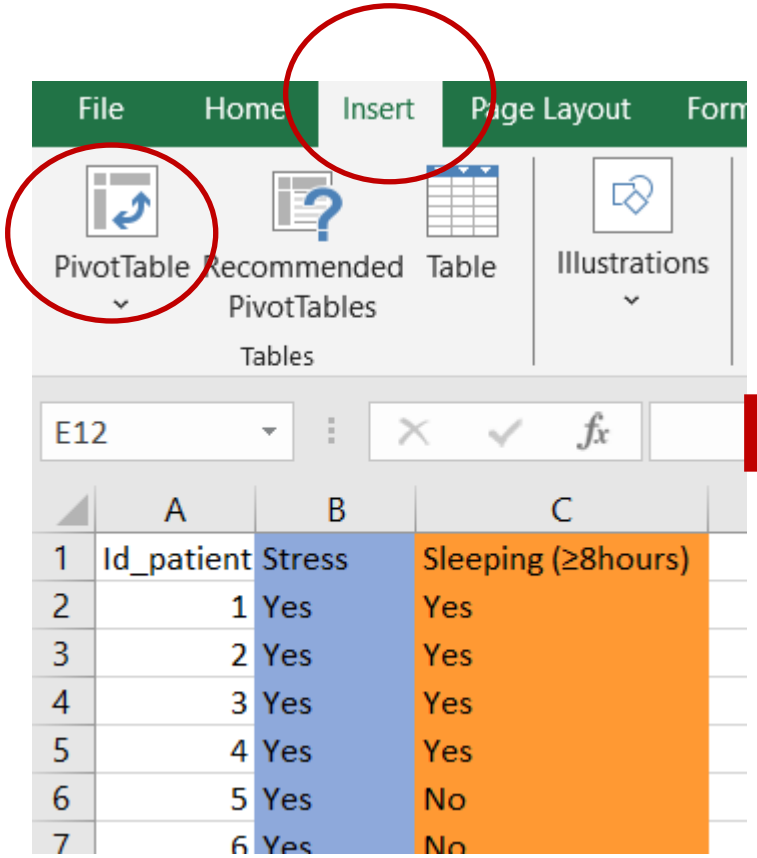
Grand total =100

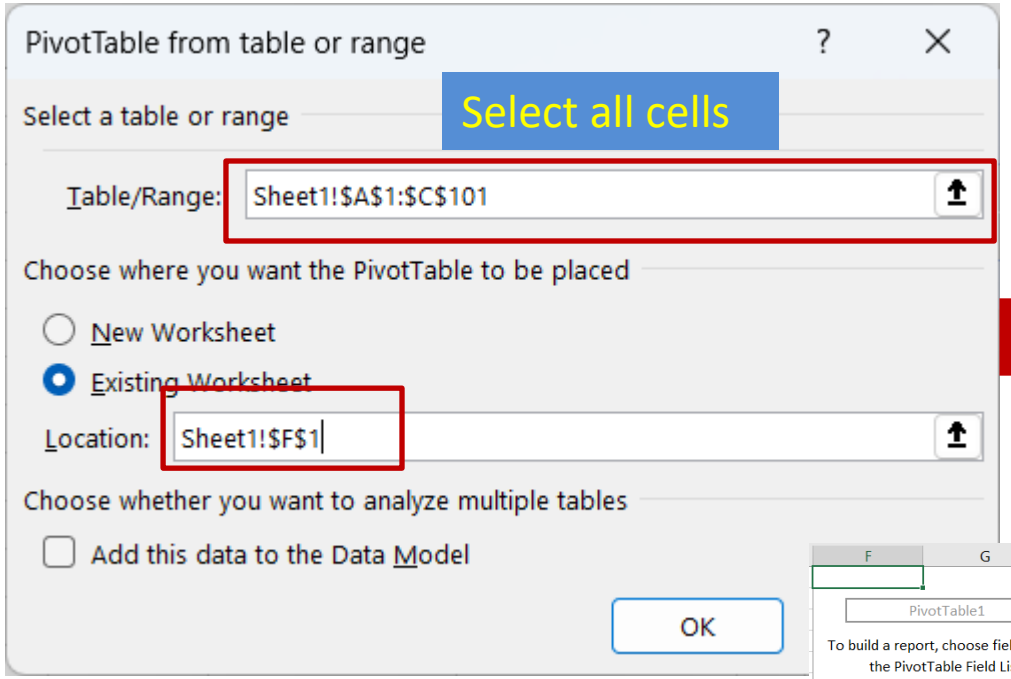
- In a study investigating the link between high sugar diet and heart disease, 400 people were investigated. 200 had a high sugar diet, 200 had a normal sugar diet. Among those with a high sugar diet, 40 had heart disease, while 160 did not. In those with a normal diet, 10 had heart disease, and 190 did not. If variable 1 is Sugar diet and variable 2 is heart disease, complete the contingency table. Consider category 1 Yes for both variables.

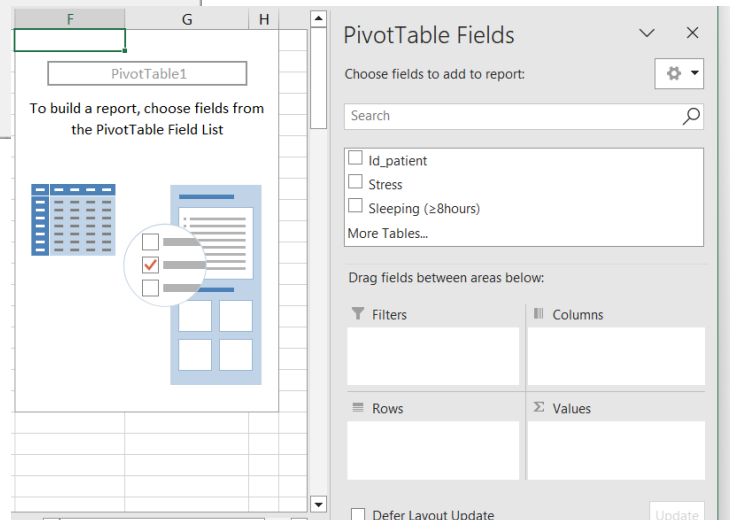
	Variable 2 – category 1	Varable 2 – category 2	Total
Variable 1 – category 1	a	b	a+b
Variable 1 – category 2	c	d	c+d
Total	a+c	b+d	n=a+b+c+d

200 had a high sugar diet, 200 had a normal sugar diet. Among those with a high sugar diet, 40 had heart disease, while 160 did not. In those with a normal diet, 10 had heart disease, and 190 did not.

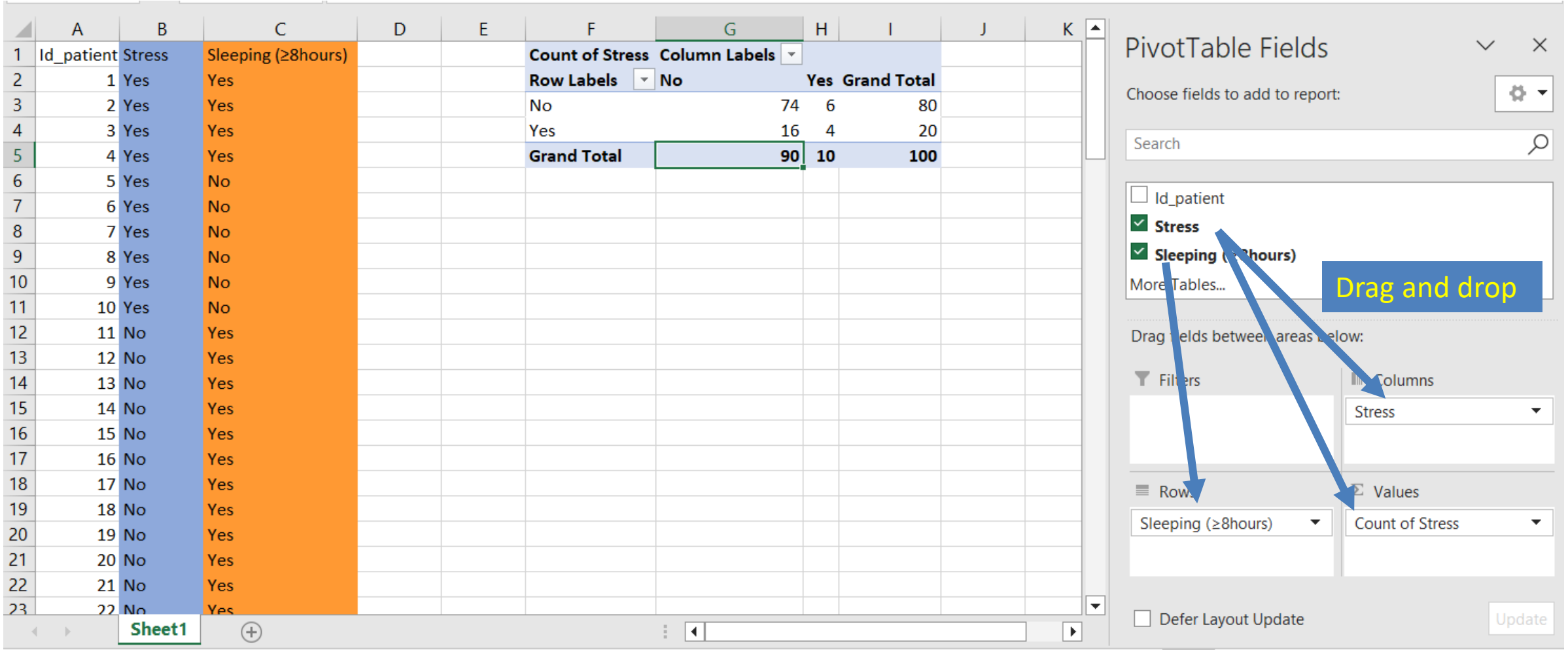
1. a=40, b=10, c=160, d=190 — 32% 23
2. a=40, b=160, c=10, d=190 — 66% 48 ✓
3. a=160, b=40, c=190, d=10 — 3% 2


File
Home
Insert
Page Layout
Form
PivotTable
Recommended PivotTables
Table
Illustrations
Tables
E12
A
B
C
1 Id_patient Stress Sleeping (≥8hours)
2 1 Yes Yes
3 2 Yes Yes
4 3 Yes Yes
5 4 Yes Yes
6 5 Yes No
7 6 Yes No


PivotTable from table or range
Select a table or range
Select all cells
Table/Range: Sheet1!$A$1:$C$101
Choose where you want the PivotTable to be placed
New Worksheet
Existing Worksheet
Location: Sheet1!$F$1
Choose whether you want to analyze multiple tables
Add this data to the Data Model
OK


PivotTable1
To build a report, choose fields from the PivotTable Field List
PivotTable Fields
Choose fields to add to report:
Search
Id_patient
Stress
Sleeping (≥8hours)
More Tables...
Drag fields between areas below:
Filters
Columns
Rows
Values
Defer Layout Update
Update

Contingency table in Excel – using Pivot table

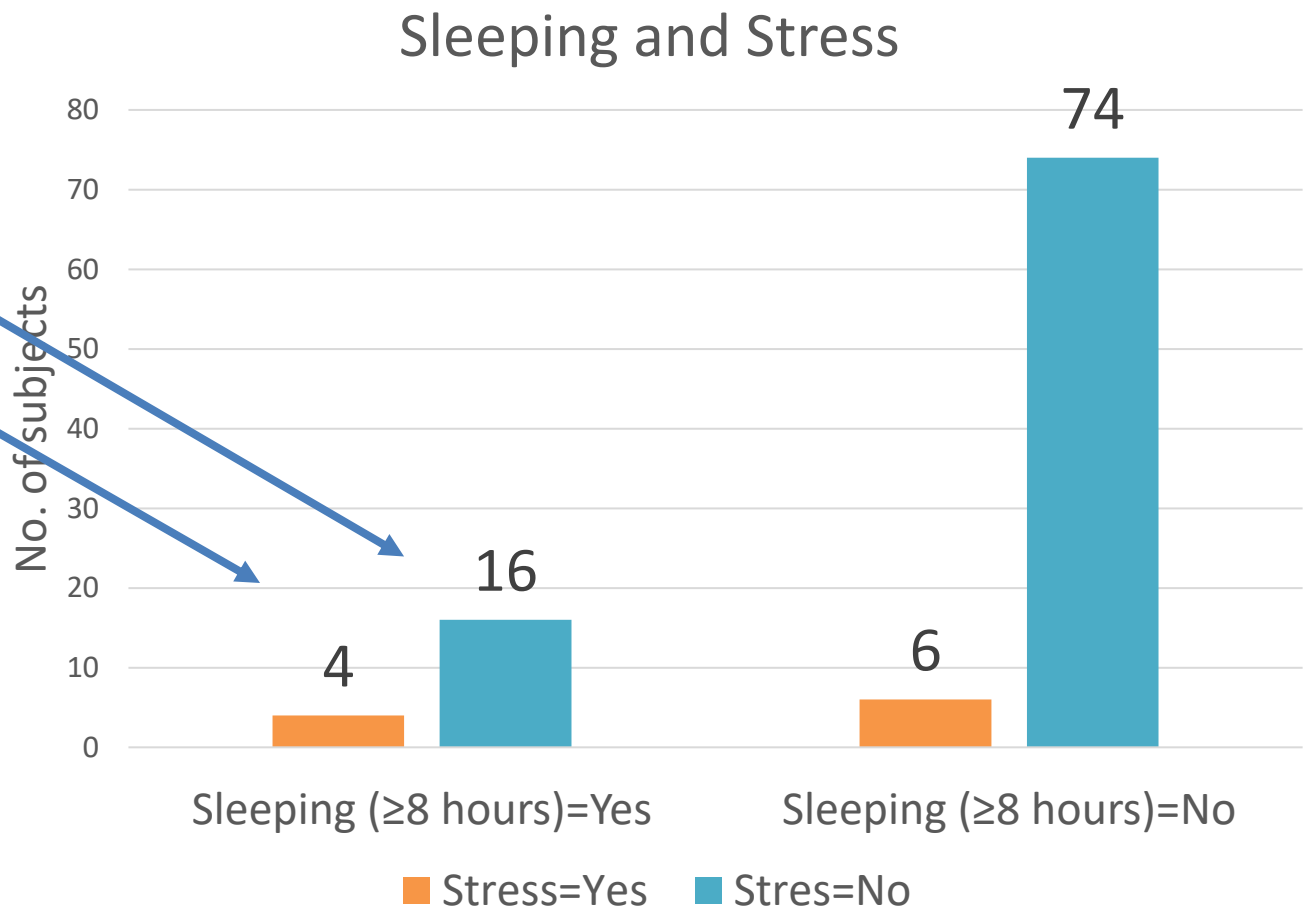


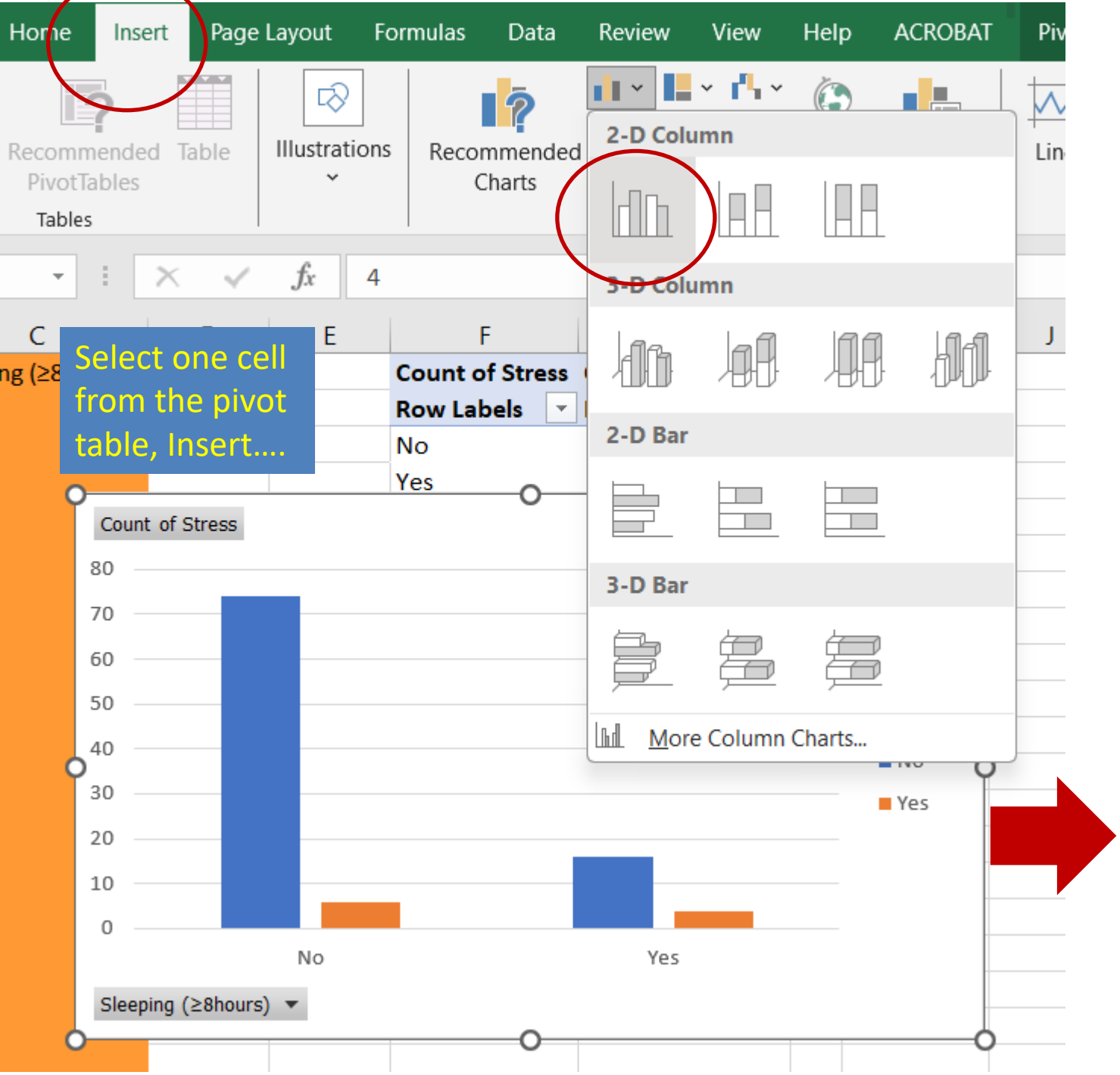
Id_patient	Stress	Sleeping (≥8hours)
1	Yes	Yes
2	Yes	Yes
3	Yes	Yes
4	Yes	Yes
5	Yes	No
6	Yes	No
7	Yes	No
8	Yes	No
9	Yes	No
10	Yes	No
11	No	Yes
12	No	Yes
13	No	Yes
14	No	Yes
15	No	Yes
16	No	Yes
17	No	Yes
18	No	Yes
19	No	Yes
20	No	Yes
21	No	Yes
22	No	Yes

Count of Stress / Row Labels	Column Labels: No	Yes	Grand Total
No	74	6	80
Yes	16	4	20
Grand Total	**90**	**10**	**100**

Chart to show the **absolute frequencies** (column chart)

	Stress Yes	Stress No	Total
Sleeping (≥8hours) Yes	4	16	20
Sleeping (≥8hours) No	6	74	80
Total	10	90	100




Home
Insert
Page Layout
Formulas
Data
Review
View
Help
ACROBAT
Recommended PivotTables
Table
Tables
Illustrations
Recommended Charts
2-D Column
3-D Column
2-D Bar
3-D Bar
More Column Charts...
Select one cell from the pivot table, Insert....
Count of Stress
Row Labels
No
Yes
Count of Stress
80
70
60
50
40
30
20
10
0
No
Yes
Yes
Sleeping (≥8hours)

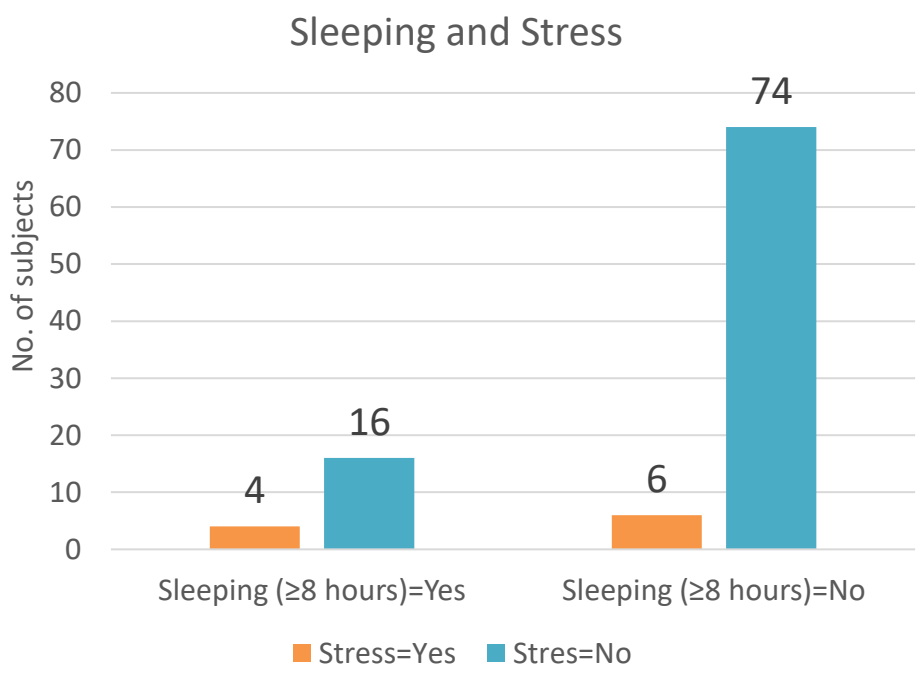
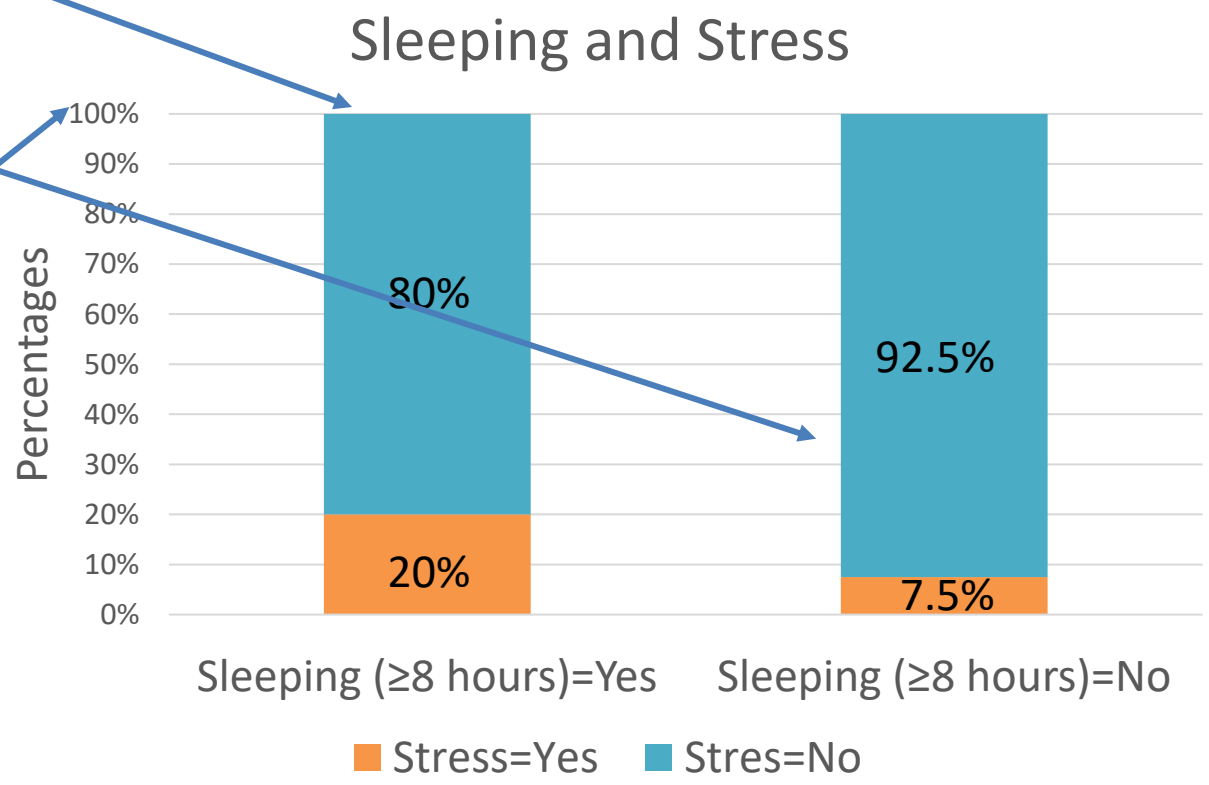
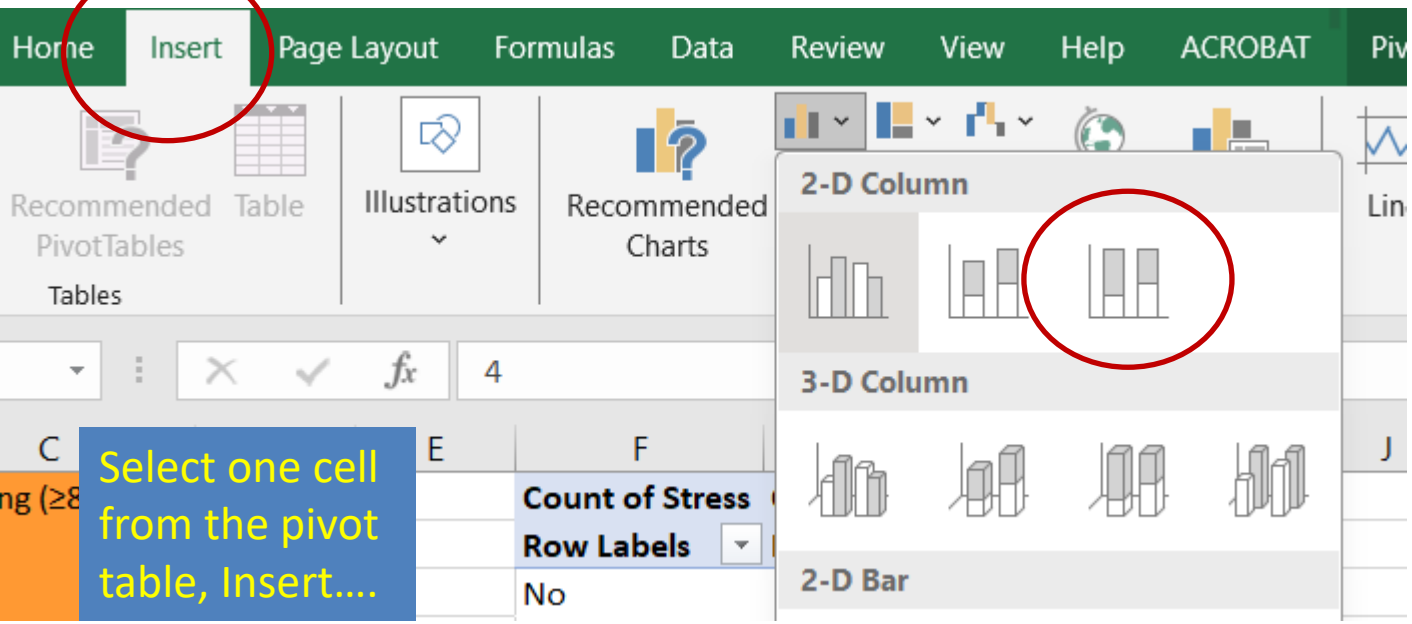

Sleeping and Stress
No. of subjects
80
70
60
50
40
30
20
10
0
4
16
6
74
Sleeping (≥8 hours)=Yes
Sleeping (≥8 hours)=No
Stress=Yes
Stres=No

Chart to show the **percentages** (relative frequencies – 100% stacked column chart)

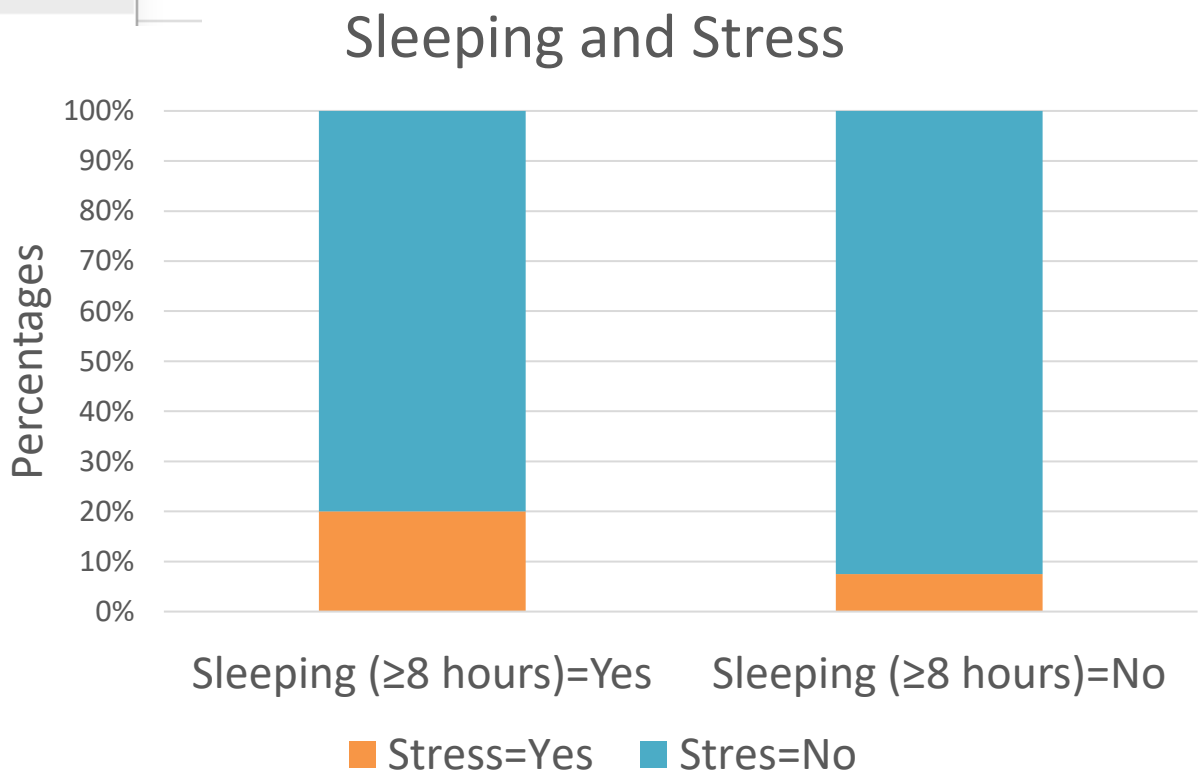
	Stress Yes	Stress No	Total
Sleeping (≥8hours) Yes	4	16	20
Sleeping (≥8hours) No	6	74	80
Total	10	90	100

- each row from the contingency table is a column representing 100%
- 20 is the total for the first column
- 80 is the total for the second column





- 100% Stacked column chart



Common Mistake

- Mixing up rows and columns

To consider that these and this are the same!

- ⚠ Watch Out! They are not!!!

	Sleeping (≥8hours) **Yes**	Sleeping (≥8hours) **No**	Total
Stress **Yes**	4	6	10
Stress **No**	16	74	80
Total	20	90	100

	Stress = **Yes**	Stress = **No**	Total
Sleeping (≥8hours) = **Yes**	4	16	20
Sleeping (≥8hours) = **No**	6	74	80
Total	10	90	100

	Stress = **No**	Stress = **Yes**	Total
Sleeping (≥8hours) = **No**	6	74	80
Sleeping (≥8hours) = **Yes**	4	16	20
Total	10	90	100

	Stress = **No**	Stress = **Yes**	Total
Sleeping (≥8hours) = **Yes**	16	4	20
Sleeping (≥8hours) = **No**	74	6	80
Total	90	10	100

Common Mistake

- Compare

	Stress = **Yes**	Stress = **No**	Total
Sleeping (≥8hours) = **Yes**	4	16	20
Sleeping (≥8hours) = **No**	6	74	80
Total	10	90	100

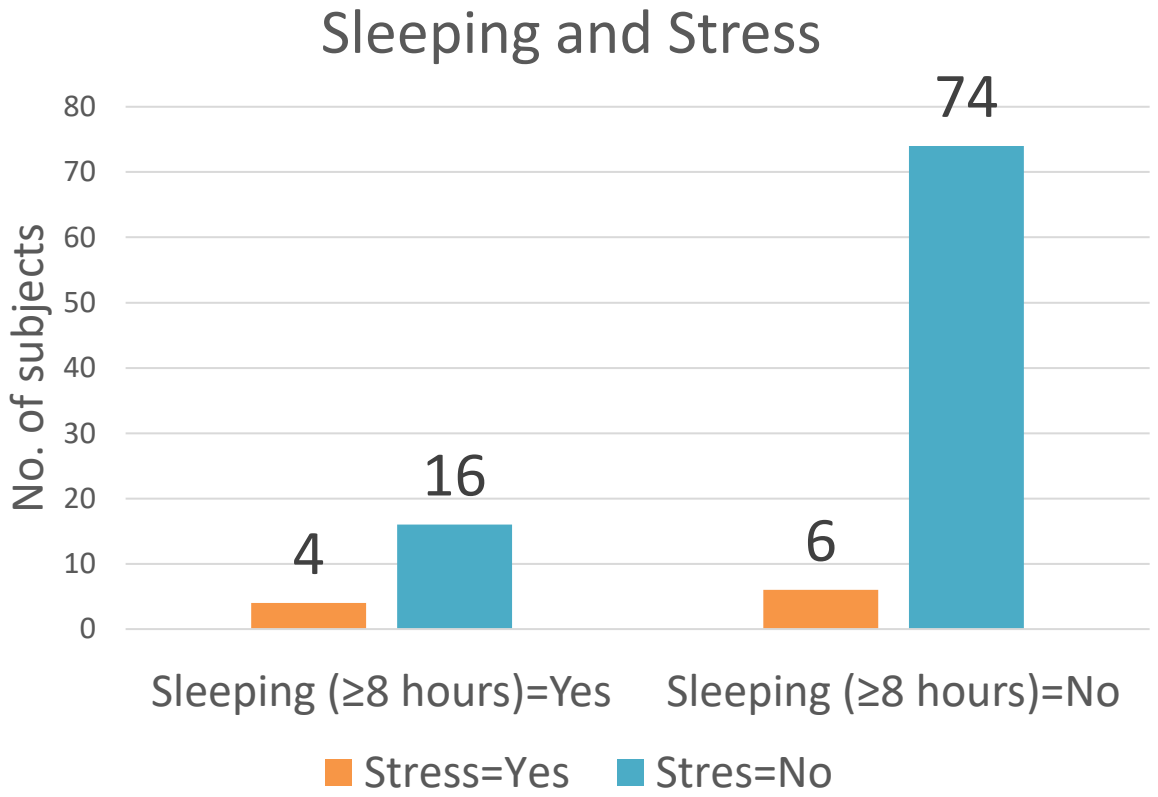
	Sleeping (≥8hours) **Yes**	Sleeping (≥8hours) **No**	Total
Stress **Yes**	4	6	10
Stress **No**	16	74	90
Total	20	80	100

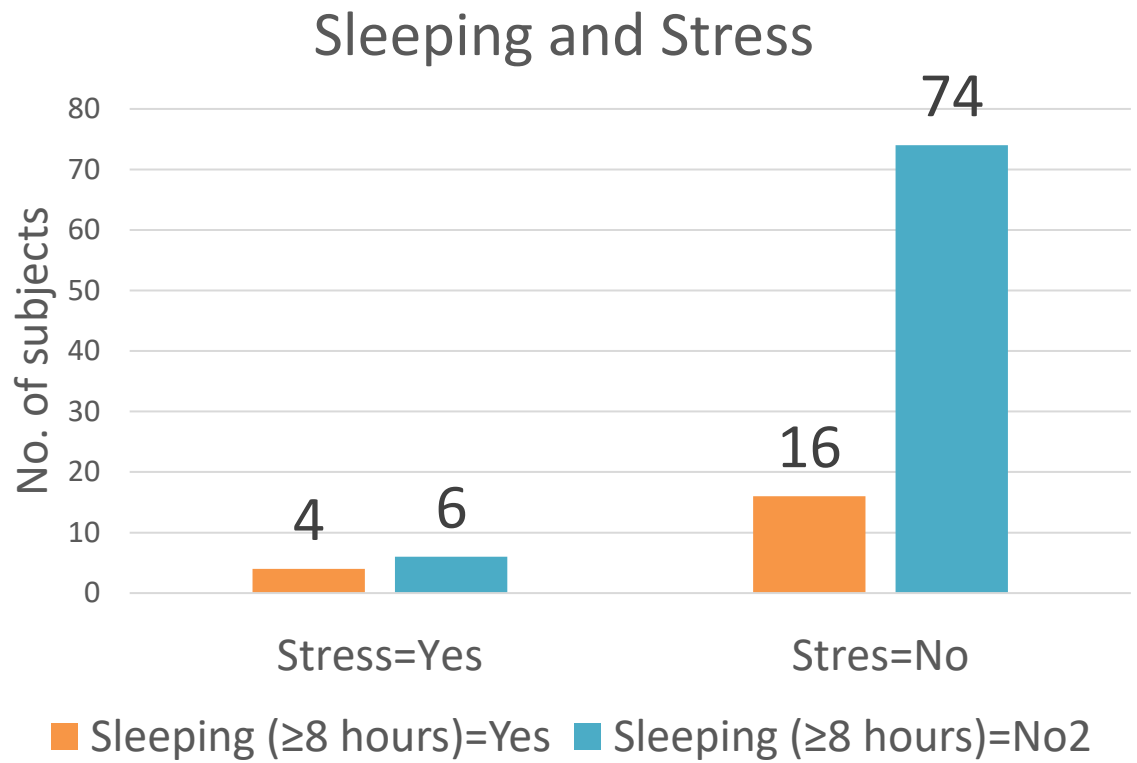
Common Mistake

- Compare

	Stress = **Yes**	Stress = **No**	Total
Sleeping (≥8hours) = **Yes**	4	16	20
Sleeping (≥8hours) = **No**	6	74	80
Total	10	90	100

	Sleeping (≥8hours) **Yes**	Sleeping (≥8hours) **No**	Total
Stress **Yes**	4	6	10
Stress **No**	16	74	90
Total	20	80	100



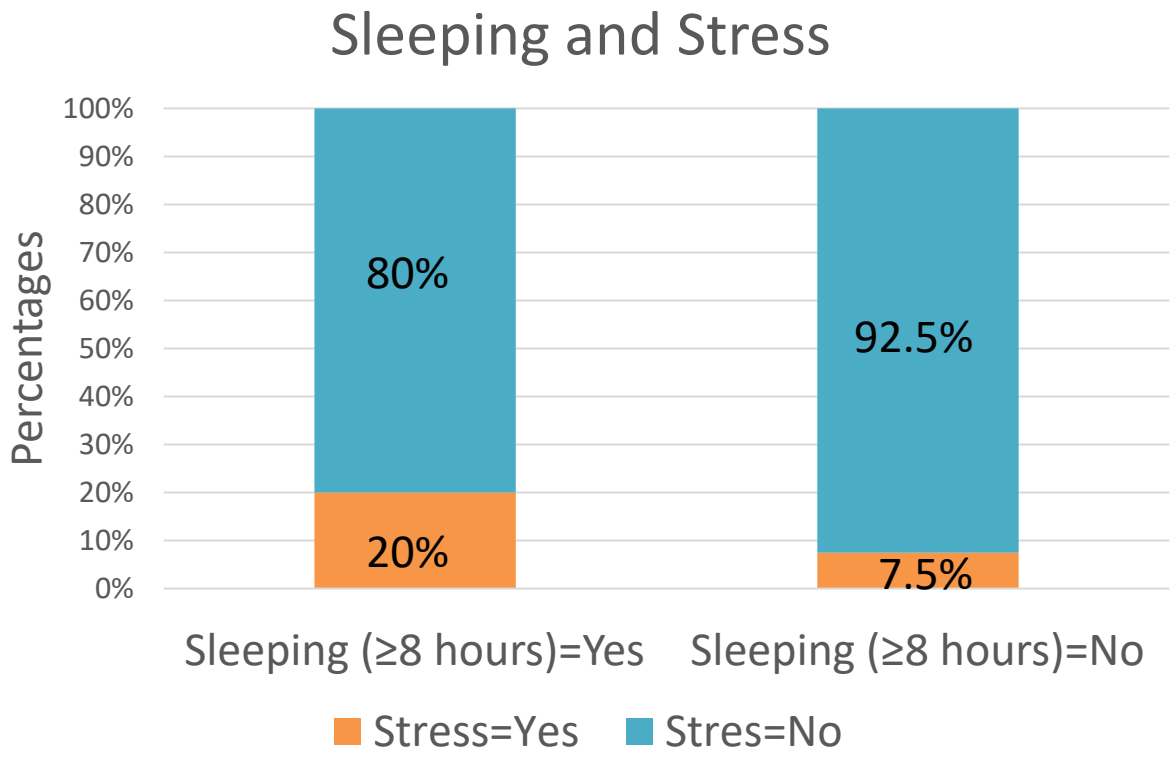


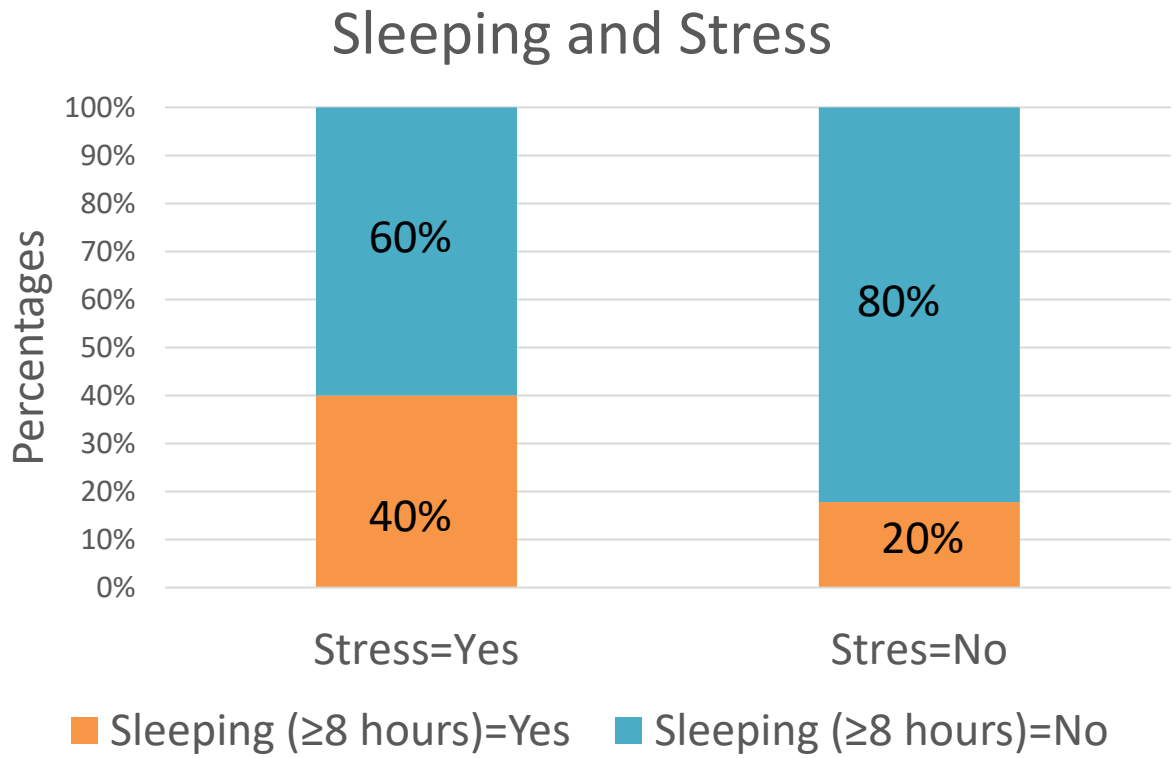
Common Mistake

The charts in % do not look the same

	Stress = **Yes**	Stress = **No**	Total
Sleeping (≥8hours) = **Yes**	4	16	20
Sleeping (≥8hours) = **No**	6	74	80
Total	10	90	100

	Sleeping (≥8hours) **Yes**	Sleeping (≥8hours) **No**	Total
Stress **Yes**	4	6	10
Stress **No**	16	74	90
Total	20	80	100





- Pay attention
  - what **a** should be!
  - what **b** should be!
  - what **c** should be!
  - what **d** should be!

	Variable 2 – category 1	Varable 2 – cathegory 1	Total
Variable 1 – category 1	a	b	a+b
Variable 1 – category 2	c	d	c+d
Total	a+c	b+d	n=a+b+c+d

Common Mistake

- Totals not adding up.

	Stress = **Yes**	Stress = **No**	Total
Sleeping (≥8hours) = **Yes**	4	16	20
Sleeping (≥8hours) = **No**	6	64	70
Total	10	80	100

Two qualitative variables

- Contingency table
  - frequency table for the association of two qualitative variables

2 dichotomial variables with 2 categories → 4 possible combined categories

2 qualitative variables with 3 categories → 9 possible combined categories

2 qualitative variables with 4 categories → 16 possible combined categories

nominal variable with 3 categories

ordinal variable with 3 categories

B	C
Hepatitis type (Type A / Type B / Type C)	**Obesity (normal weight/overweight /obese)**
Type C	Normal weight
Type C	Normal weight
Type C	Normal weight
Type B	Overweight
Type C	Normal weight
Type C	Normal weight
Type C	Normal weight
Type B	Overweight
Type A	Overweight
Type A	Overweight
Type B	Obese
Type A	Obese

…

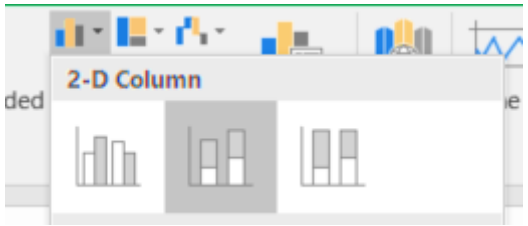
categories are in order

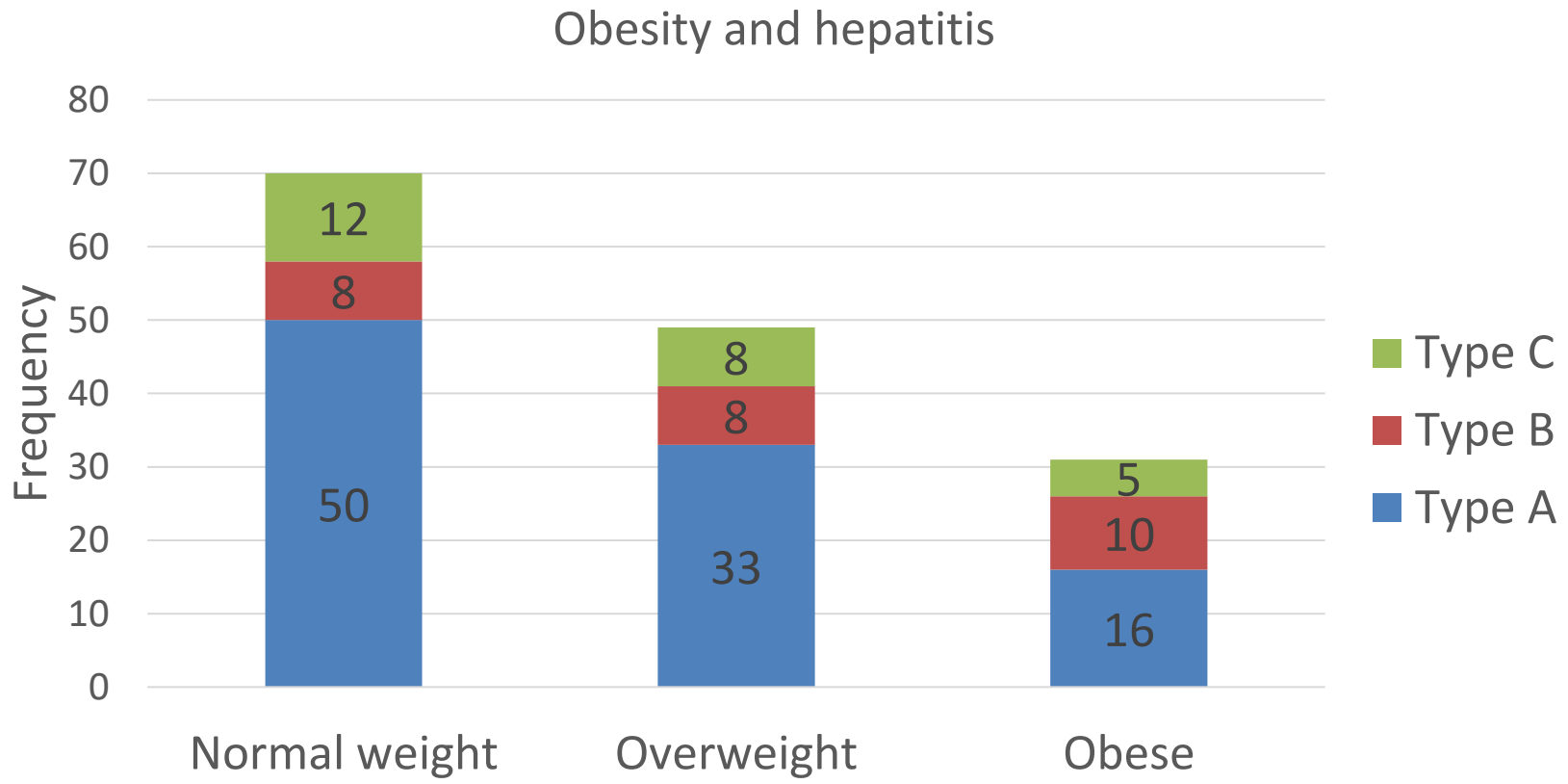
Contingency table

	Hepatitis type			
Body weight	Type A	Type B	Type C	Grand Total
Normal weight	50	8	12	70
Overweight	33	8	8	49
Obese	16	10	5	31
Grand Total	99	26	25	150

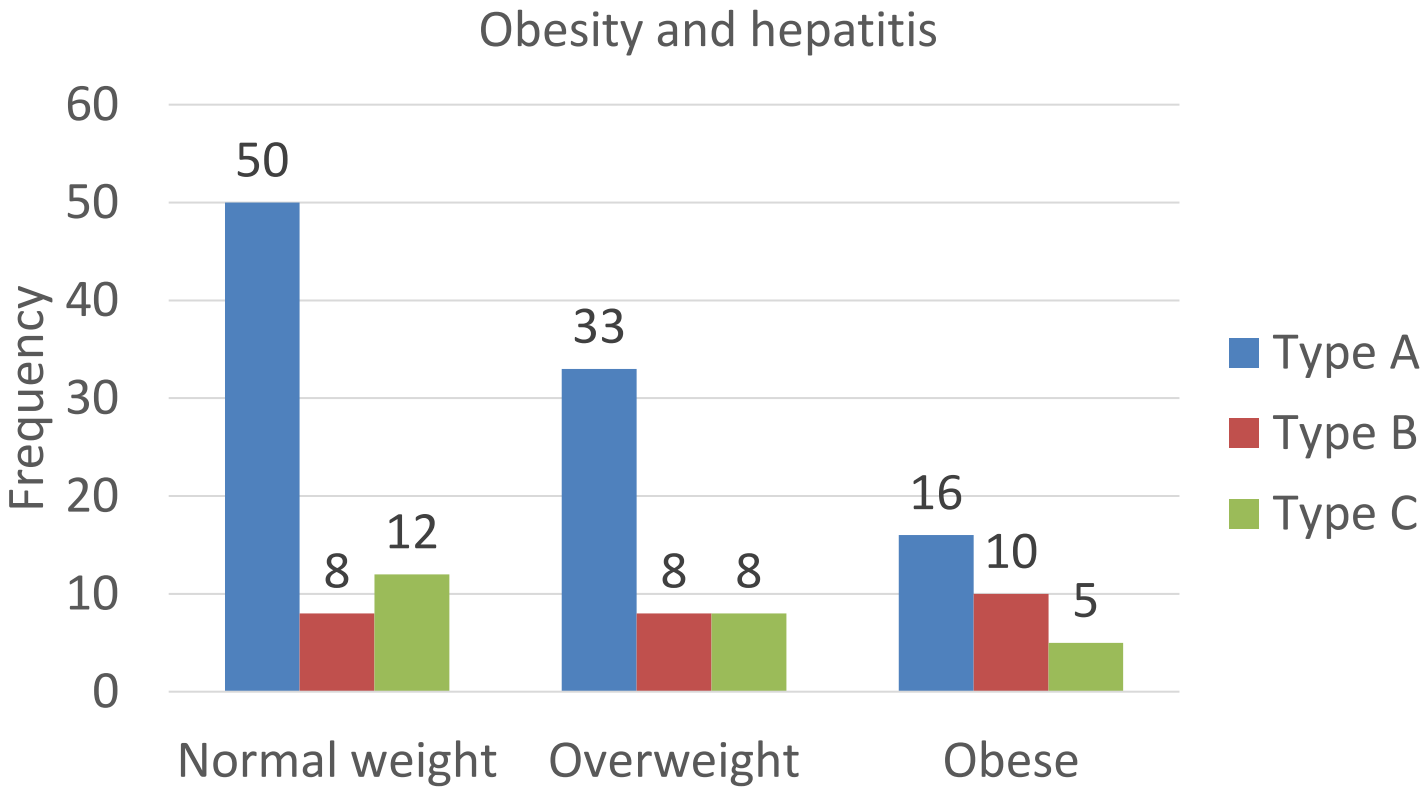
	Hepatitis type			
Body weight	Type A	Type B	Type C	Grand Total
Normal weight	50	8	12	70
Overweight	33	8	8	49
Obese	16	10	5	31
Grand Total	99	26	25	150

Stacked column chart

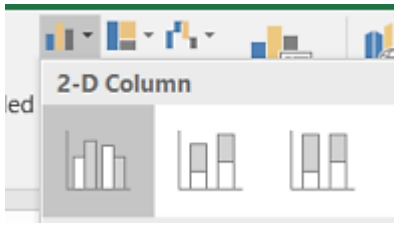




	Hepatitis type			
Body weight	Type A	Type B	Type C	Grand Total
Normal weight	50	8	12	70
Overweight	33	8	8	49
Obese	16	10	5	31
Grand Total	99	26	25	150

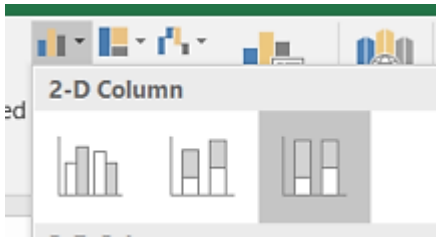


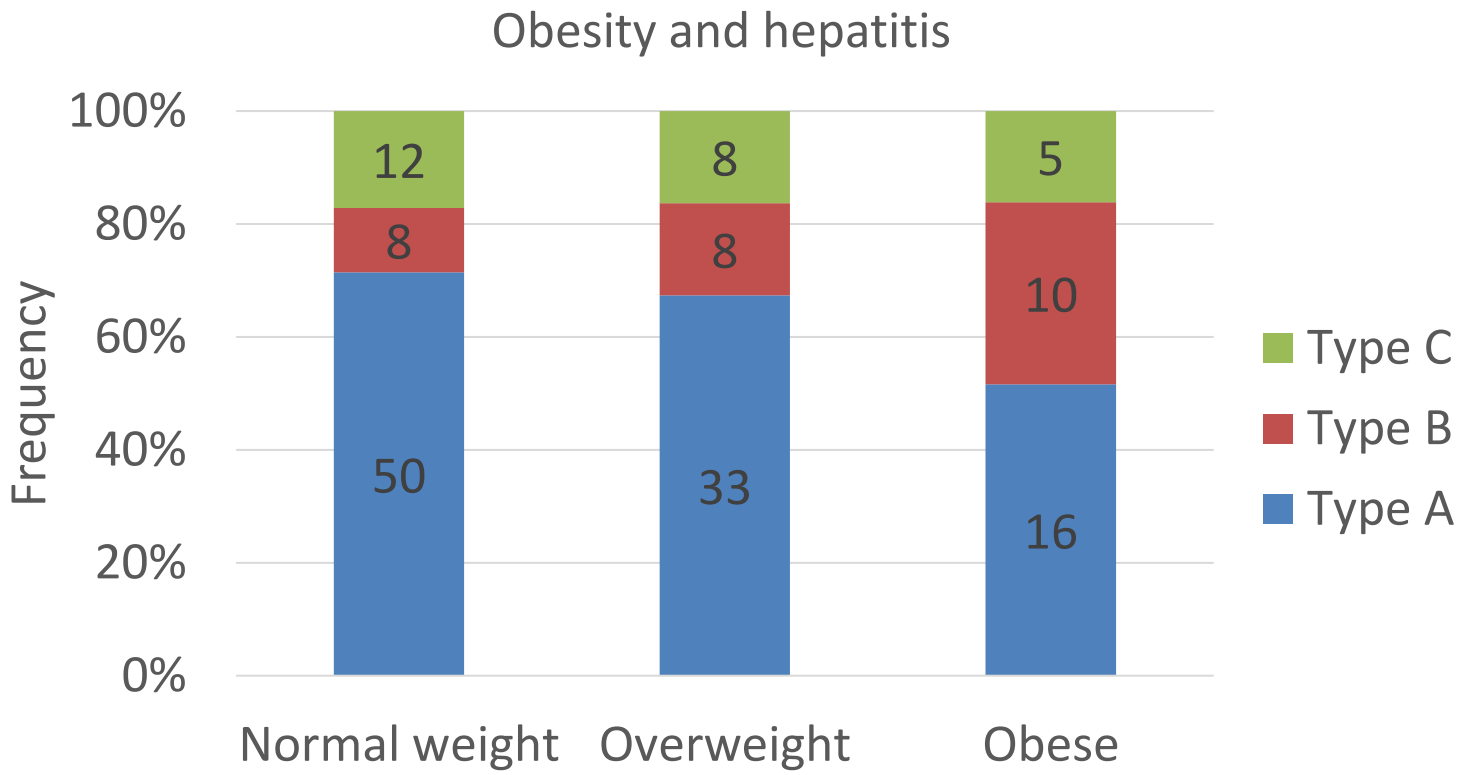
Column chart



	Hepatitis type			
Body weight	Type A	Type B	Type C	Grand Total
Normal weight	50	8	12	70
Overweight	33	8	8	49
Obese	16	10	5	31
Grand Total	99	26	25	150

100% Stacked column chart





- Contingency table

1 qualitative variable with 2 categories

1 qualitative variable with 3 categories

6 possible combined categories

Generalization

1 qualitative variable with n categories

1 qualitative variable with m categories

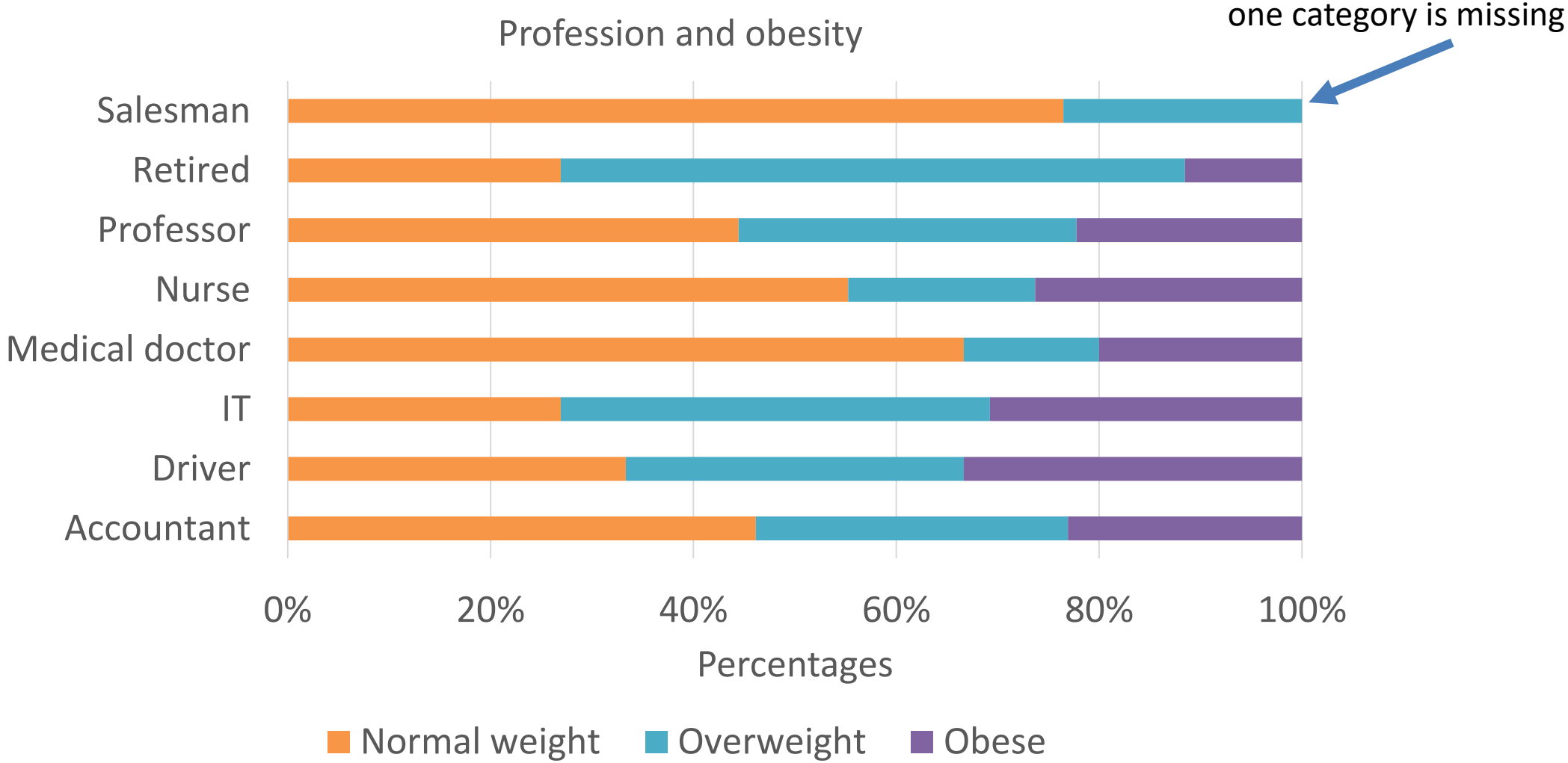
mxn possible combined categories

1 qualitative variable with 8 categories

1 qualitative variable with 3 categories

maximum 8x3=24 possible combined categories

one category is missing



Summarizing qualitative variables representations

One qualitative variable

- Frequencies Tables

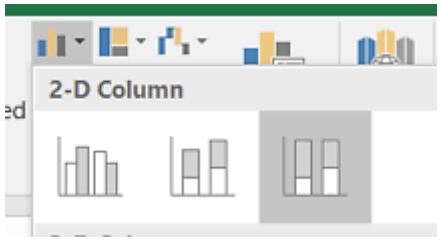
- Charts
  - Pie
  - Column
  - Bar

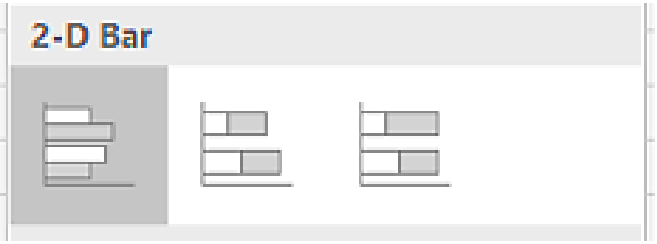
Two qualitative variables

- Contingency table

Charts

- Column / Bar
- Stacked Column / Bar
- 100% Stacked Column / Bar





Objectives

- Summarizing **ORDINAL** qualitative variables
  - Frequencies
  - Tables
  - Charts
  - Descriptive statistics parameters

Question 1
Somewhat disagree
Strongly disagree
Somewhat agree
Somewhat agree
Somewhat disagree
Somewhat disagree
Strongly agree
Strongly agree
Strongly agree
Neither agree or disagree
Somewhat agree
Somewhat disagree
Somewhat disagree
Neither agree or disagree
Neither agree or disagree
Somewhat agree
Somewhat agree
Somewhat disagree
Somewhat disagree
Neither agree or disagree

C
Glycemia classification (hypoglicemia, normal glycemia, prediabetes, diabetes)
normal glycemia
normal glycemia
normal glycemia
normal glycemia
prediabetes
diabetes
diabetes
prediabetes
normal glycemia
normal glycemia
normal glycemia
normal glycemia
normal glycemia
prediabetes
prediabetes
diabetes
normal glycemia
normal glycemia
normal glycemia
hypoglicemia

Apgar Score
10
10
9
10
10
8
8
9
9
10
10
10
10
6
10
10
10
7
7
8

C
MELD score
5
1
6
9
11
8
25
30
32
14
8
7
6
5
22
25
29
12
16
19

Ordinal variables

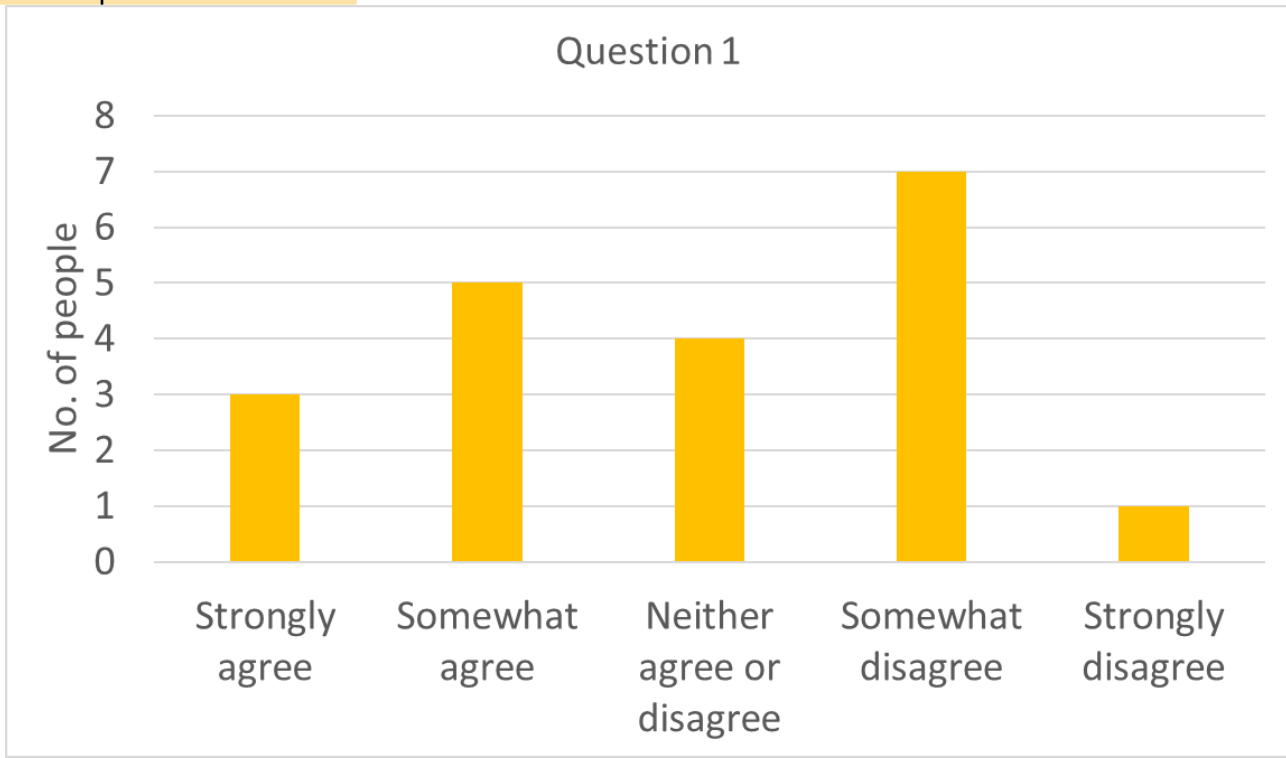
- not scores
  - transformed from quantitative variables
- scores with few values
- scores with many values

Model of End-Stage Liver Disease (MELD) score – the need to liver transplant
>20 less than 90 days of survival

Qualitative variable - Column/Bar Chart

	Question 1
1	Question 1
2	Somewhat disagree
3	Strongly disagree
4	Somewhat agree
5	Somewhat agree
6	Somewhat disagree
7	Somewhat disagree
8	Strongly agree
9	Strongly agree
10	Strongly agree
11	Neither agree or disagree
12	Somewhat agree
13	Somewhat disagree
14	Somewhat disagree
15	Neither agree or disagree
16	Neither agree or disagree
17	Somewhat agree
18	Somewhat agree
19	Somewhat disagree
20	Somewhat disagree
21	Neither agree or disagree

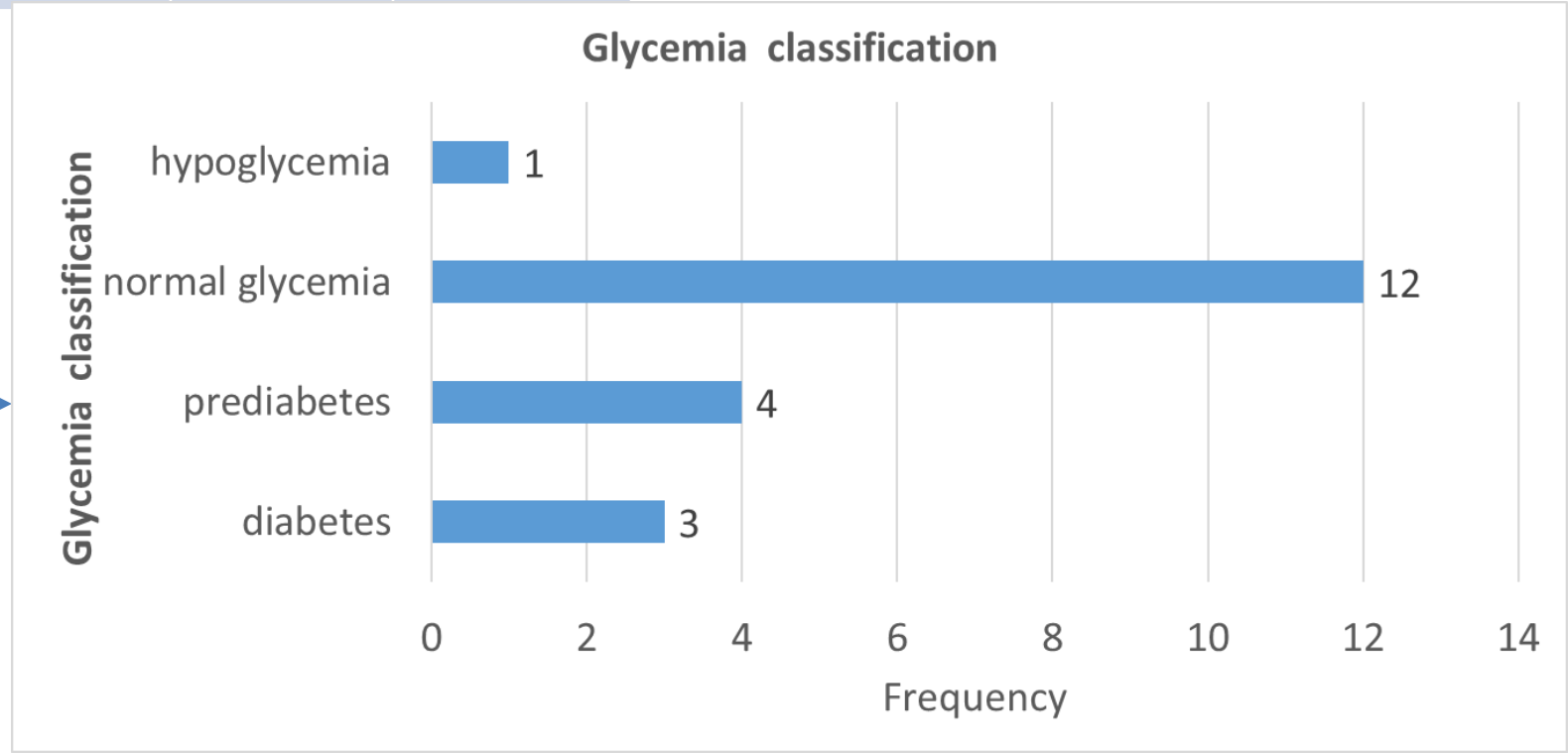
Question 1	Absolute frequency	Relative frequency
Strongly agree	3	15
Somewhat agree	5	25
Neither agree or disagree	4	20
Somewhat disagree	7	35
Strongly disagree	1	5
Total	20	100



Transformed variable from a quantitative variable – Column/Bar

C
Glycemia classification (hypoglicemia, normal glycemia, prediabetes, diabetes)
normal glycemia
normal glycemia
normal glycemia
normal glycemia
prediabetes
diabetes
diabetes
prediabetes
normal glycemia
normal glycemia
normal glycemia
normal glycemia
normal glycemia
prediabetes
prediabetes
diabetes
normal glycemia
normal glycemia
normal glycemia
hypoglicemia

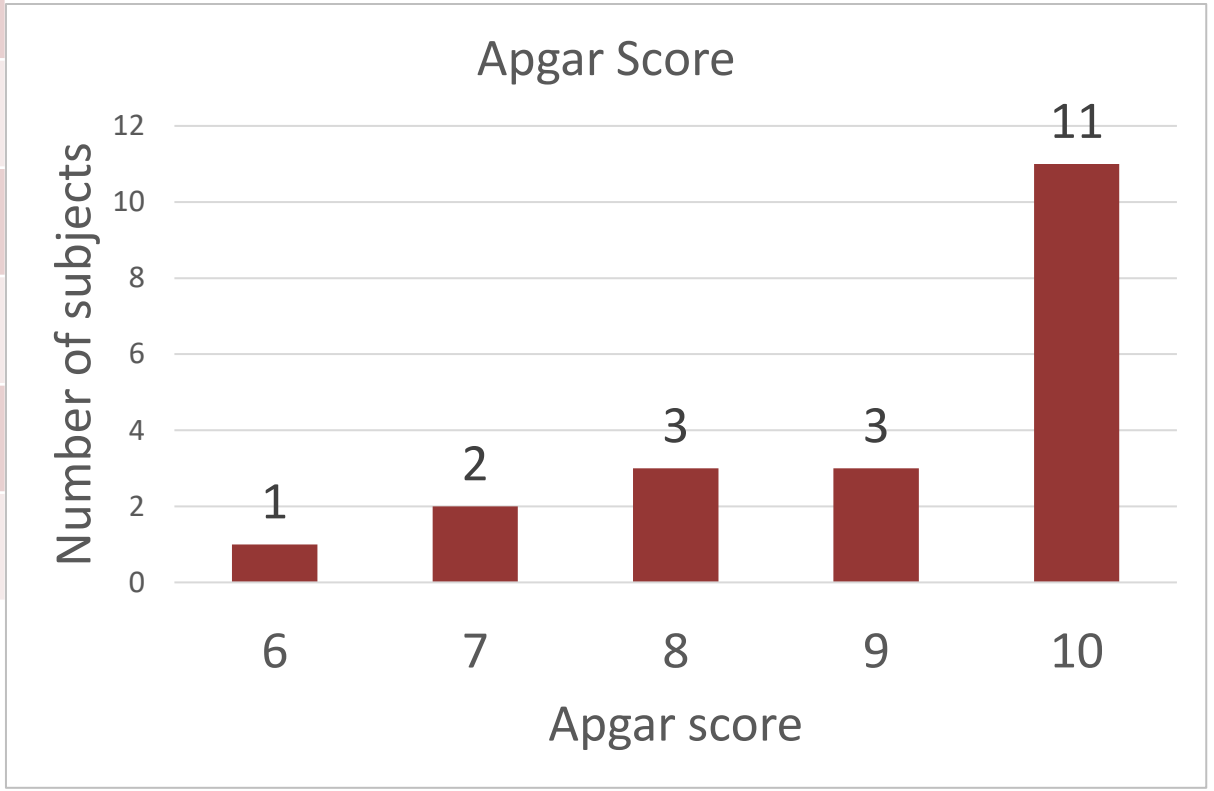
Glycemia classification (hypoglicemia, normal glycemia, prediabetes, diabetes)	Absolute frequency	Relative frequency
diabetes	3	15
hypoglycemia	4	20
normal glycemia	12	60
prediabetes	1	5
Total	20	100



Apgar Score
10
10
9
10
10
8
8
9
9
10
10
10
10
10
6
10
10
10
7
7
8

Chart – Column/Bar Chart

Apgar score	Absolute frequency	Relative frequency (%)
10	11	55
9	3	15
8	3	15
7	2	10
6	1	5
Total	20	100



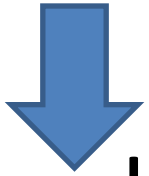
C
MELD score
5
1
6
9
11
8
25
30
32
14
8
7
6
5
22
25
29
12
16
19

MELD score (0-32)	Frequency
0	0
1	1
2	1
3	0
4	0
5	1
6	2
7	1
8	2
9	1
10	0
11	1
12	1
13	0
14	1
15	1
16	1

MELD score (0-32)	Frequency
17	0
18	0
19	1
20	0
21	0
22	1
23	0
24	0
25	1
26	0
27	0
28	0
29	1
30	1
31	0
32	1

Scores with many values

If represent it as usuall



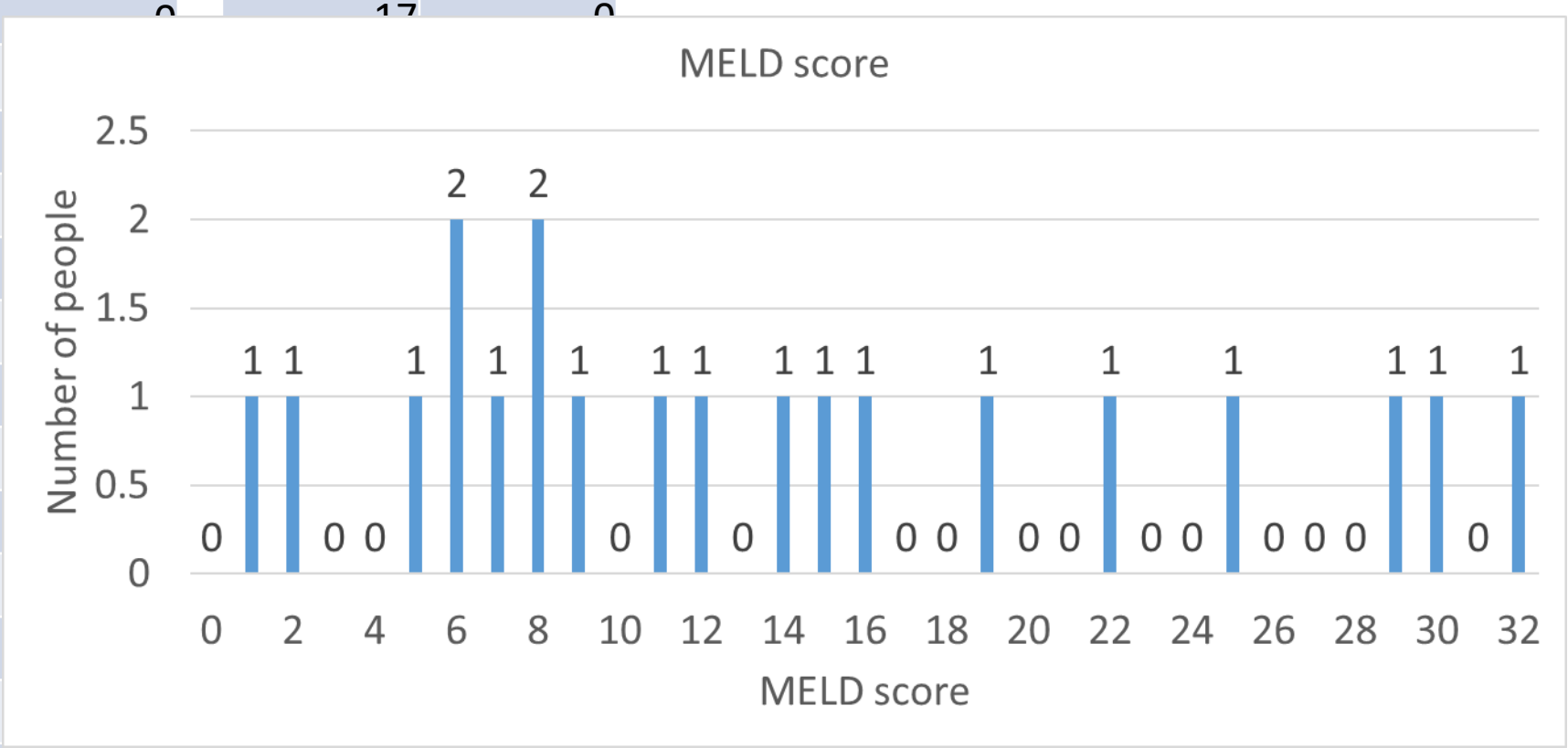
A very long frequency table

Model of End-Stage Liver Disease (MELD) score – the need to liver transplant >20 less than 90 days of survival

C
MELD score
5
1
6
9
11
8
25
30
32
14
8
7
6
5
22
25
29
12
16
19

MELD score	Frequency
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	1
13	0
14	1
15	1
16	1

MELD score	Frequency
	0
29	1
30	1
31	0
32	1



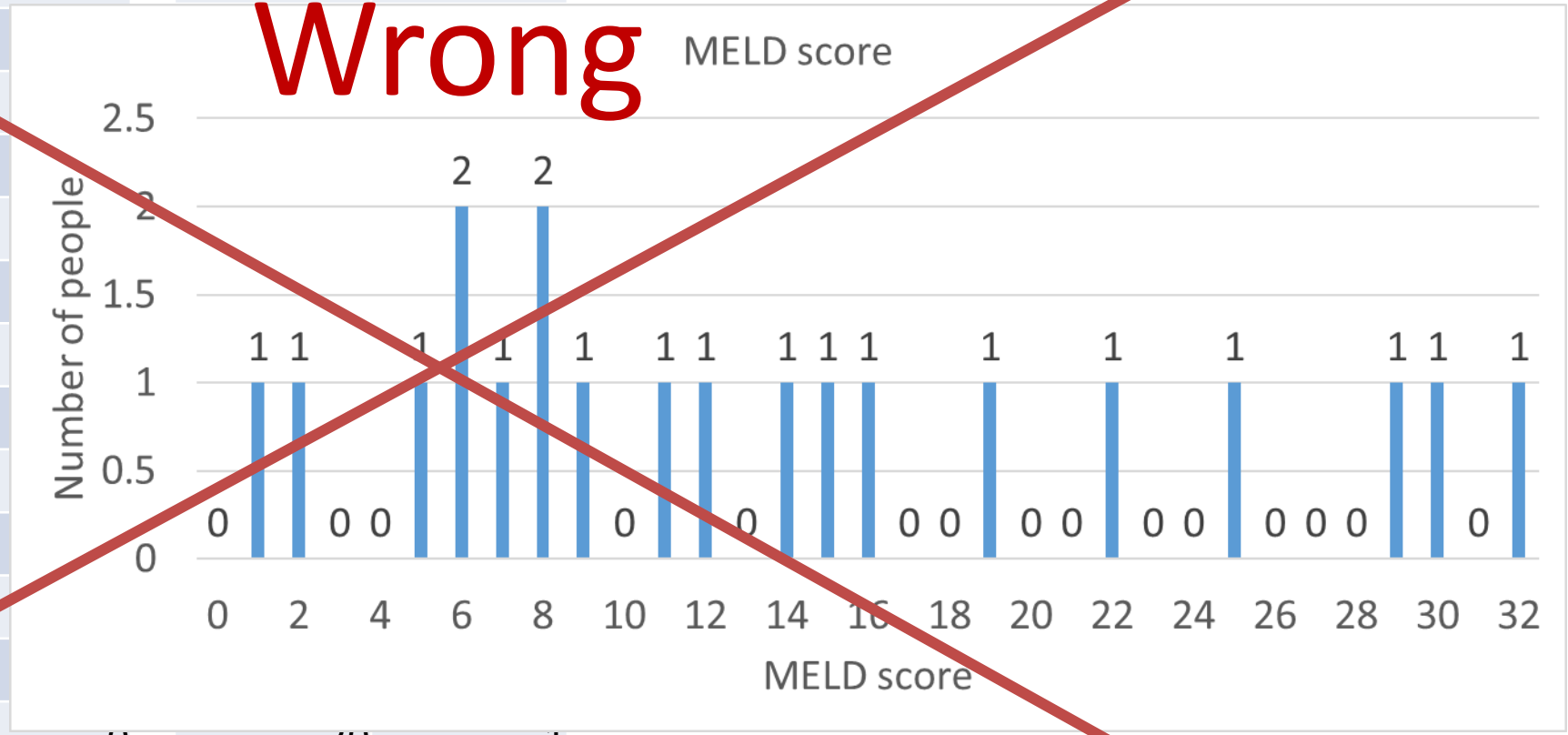
an incomprehensible and not usefull graph

C
MELD score
5
1
6
9
11
8
25
30
32
14
8
7
6
5
22
25
29
12
16
19

MELD score	Frequency
0	0
1	1
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	0
14	1
15	1
16	1

MELD score	Frequency
17	0
18	0
30	1
31	0
32	1

Wrong



C
MELD score
5
1
6
9
11
8
25
30
32
14
8
7
6
5
22
25
29
12
16
19

MELD score (0-32)
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

MELD score (0-32)
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

MELD scores
less or equal than 8
from 9 to 16
from 17 to 24
from 25 to 32

transform: few categories

- divide it in equal intervals
- from 8 to 8
- first or last interval can be grater than the others

C
MELD score
5
1
6
9
11
8
25
30
32
14
8
7
6
5
22
25
29
12
16
19

MELD scores (0-32)	Absolute frequency
0	0
1	1
2	1
3	0
4	0
5	1
6	2
7	1
8	2
9	1
10	0
11	1
12	1
13	0
14	1
15	1
16	1

MELD scores (0-32)	Absolute frequency
17	0
18	0
19	1
20	0
21	0
22	1
23	0
24	0
25	1
26	0
27	0
28	0
29	1
30	1
31	0
32	1

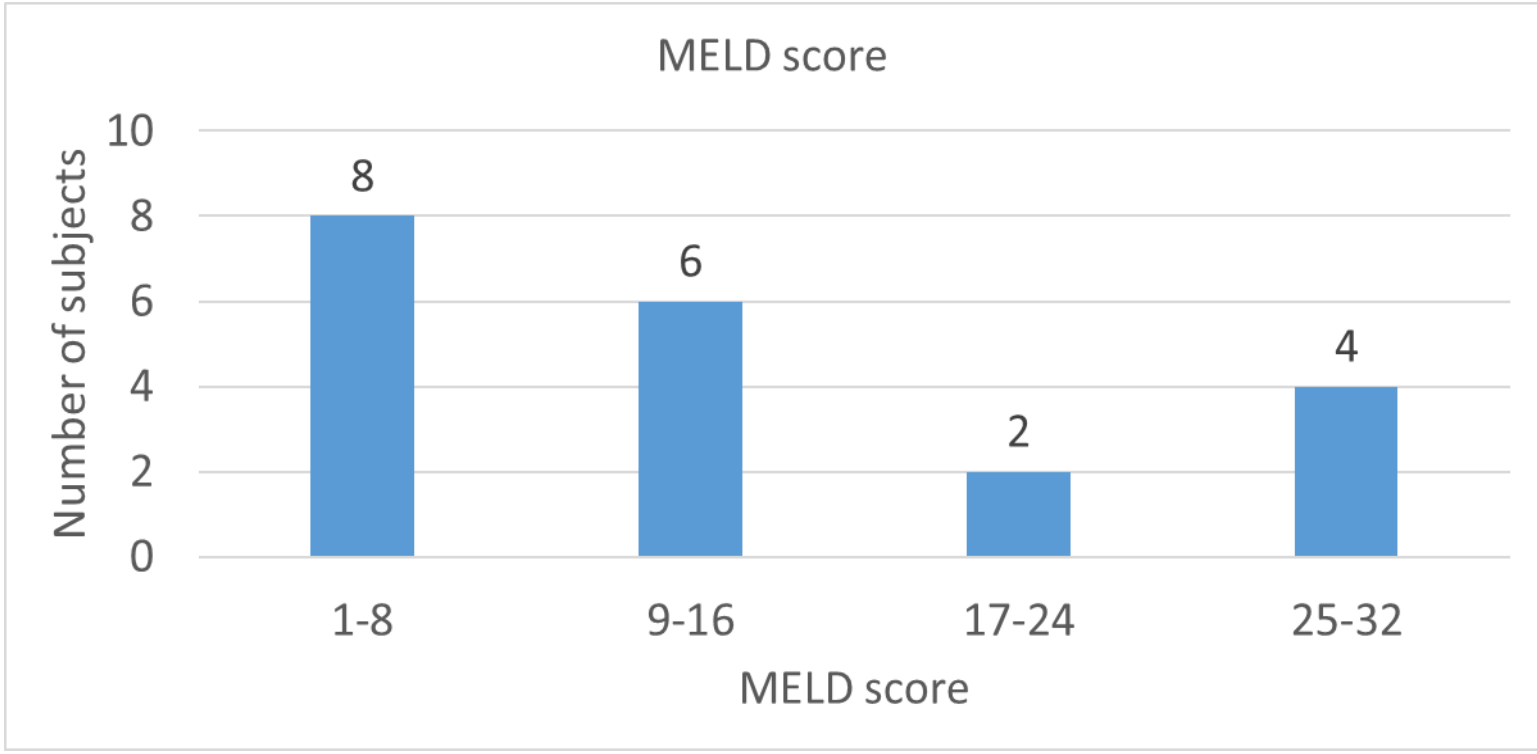
Count how many data are in each interval

MELD scores	Absolute frequency
less or equal than 8	8
from 9 to 16	6
from 17 to 24	2
from 25 to 32	4
Total	20

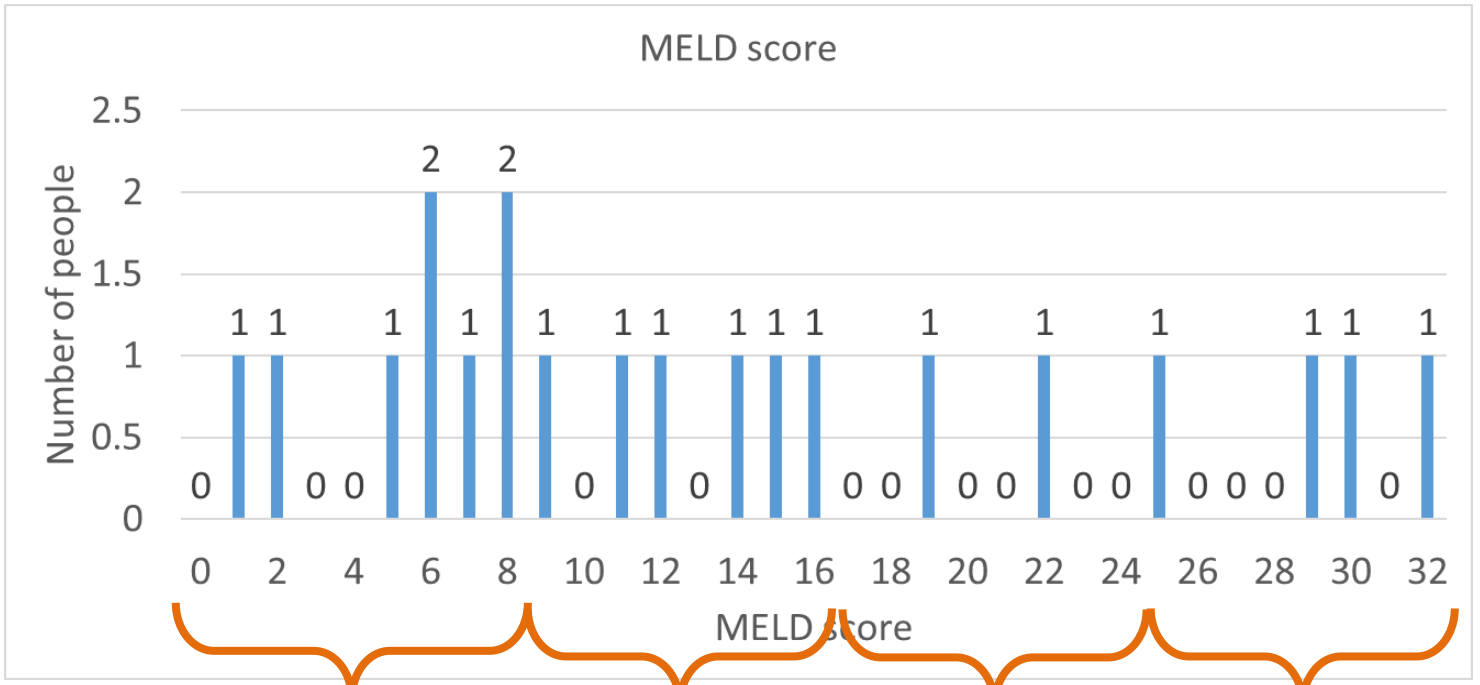
C
MELD score
5
1
6
9
11
8
25
30
32
14
8
7
6
5
22
25
29
12
16
19

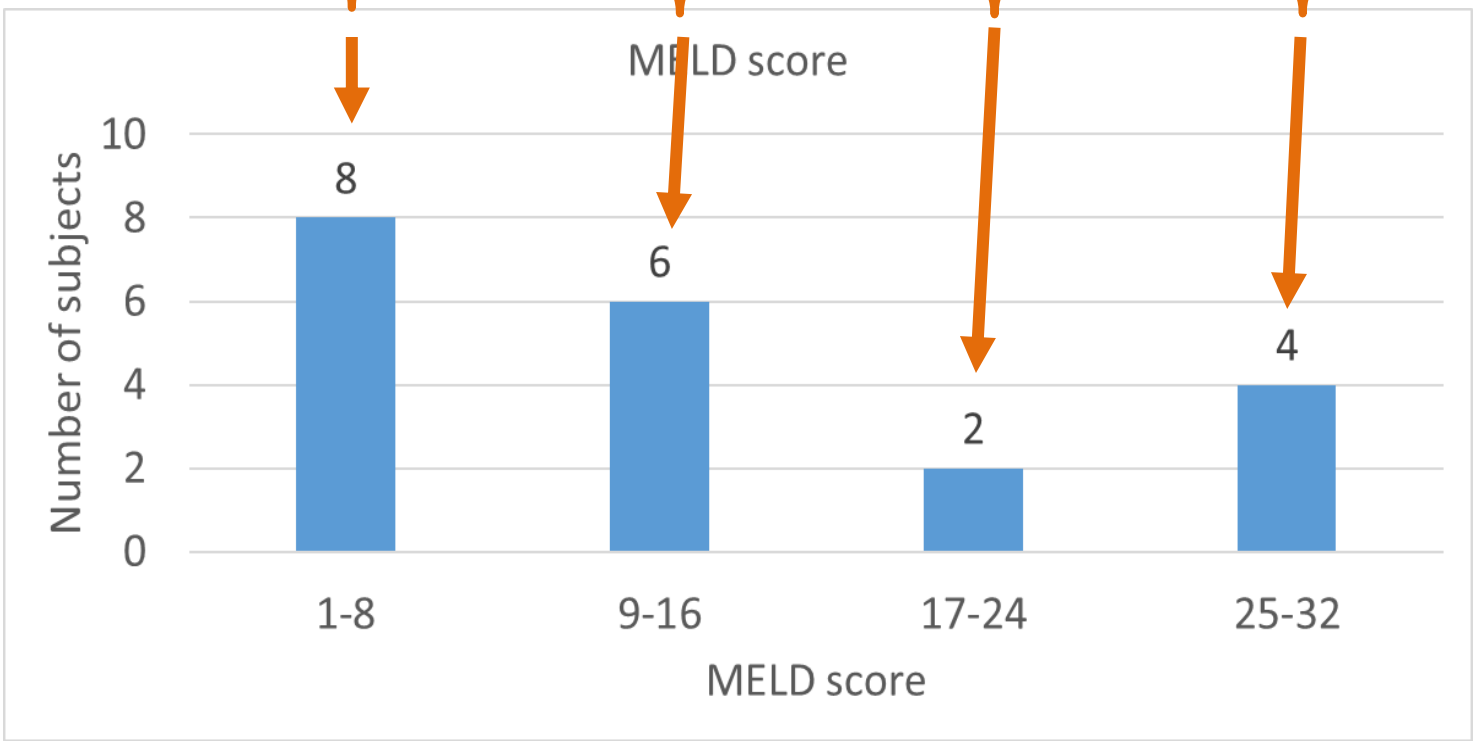
MELD scores	Absolute frequency	Relative frequency
less or equal than 8	8	40
from 9 to 16	6	30
from 17 to 24	2	10
from 25 to 32	4	20
Total	20	100

Column or bar chart



C
MELD score
5
1
6
9
11
8
25
30
32
14
8
7
6
5
22
25
29
12
16
19

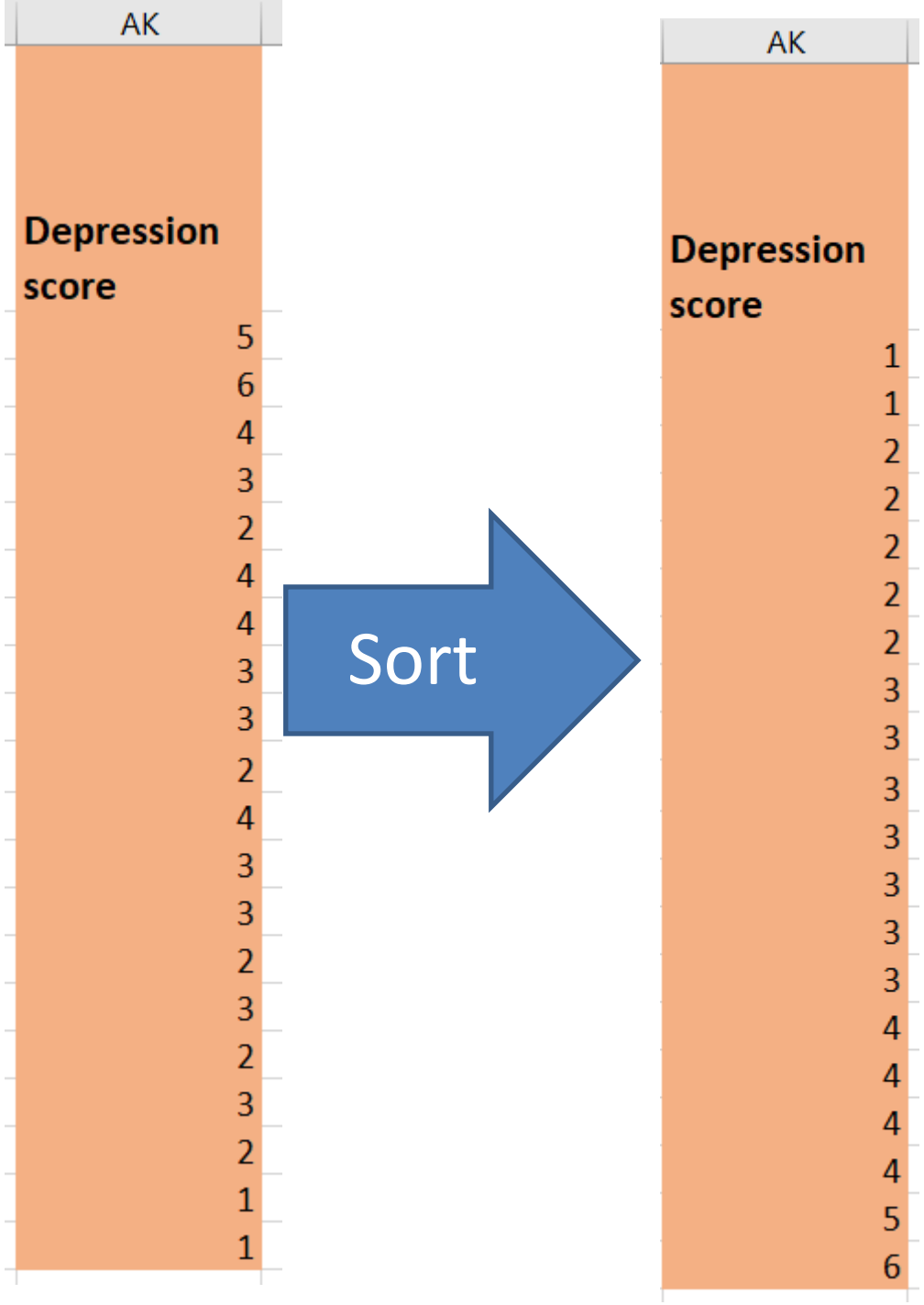




We can compare the categories

- the most frequent interval is 0 to 8
- the leet frequent interval is 17 to 24

Descriptive statistics parameters for ordinal variables



Measures for ordinal variables: scores

Minimum = 1

Maximum = 6

Range = Maximum – Minimum =

= 6 - 1 = 5

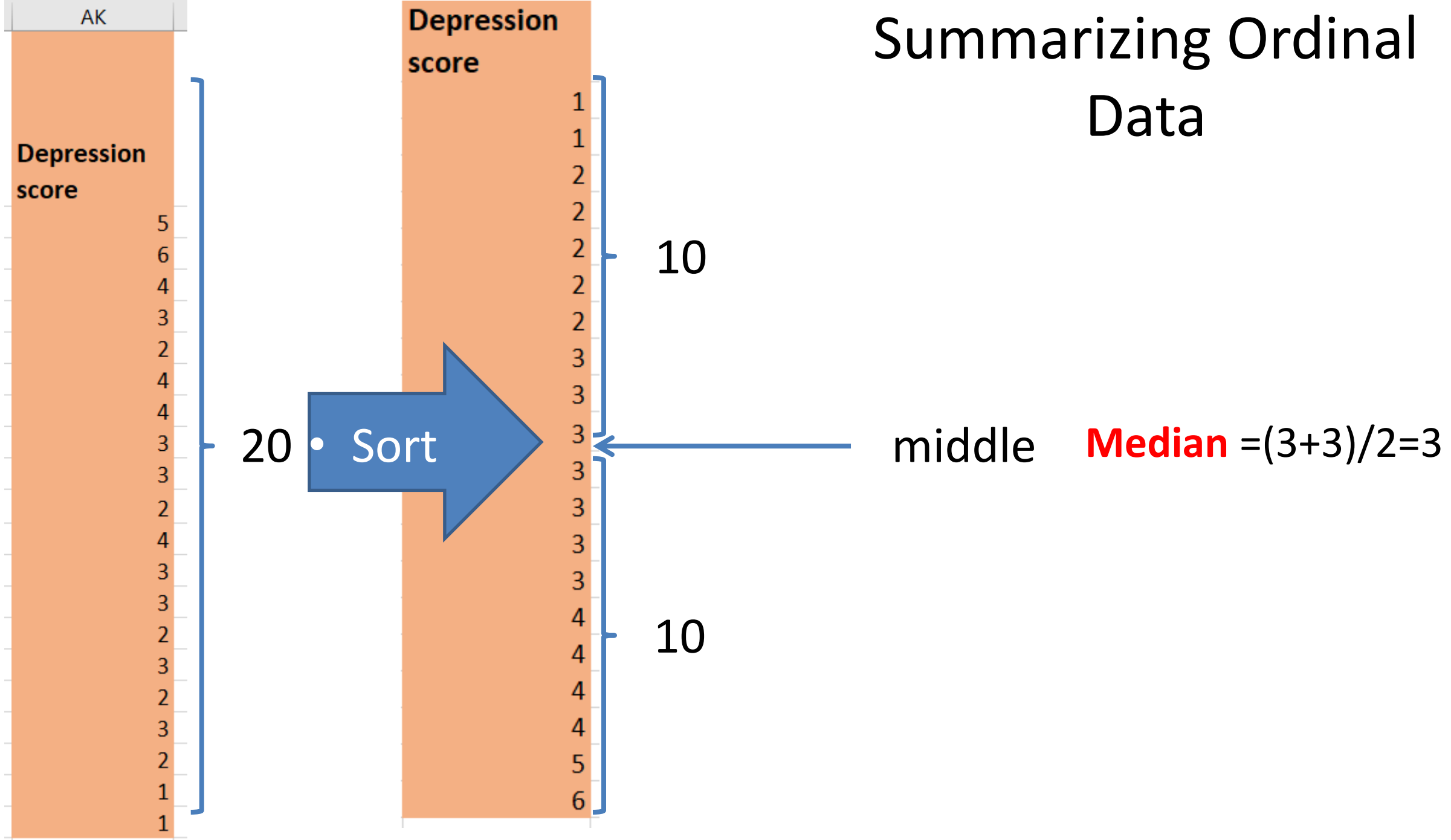
Summarizing Ordinal Data

Minimum – the smallest score

Maximum – the biggest score

Range – the difference between maximum and minimum score

Summarizing Ordinal Data



Summarizing Ordinal Data

Median – the middle of observation,

the point at which half the observations are smaller and half are larger.

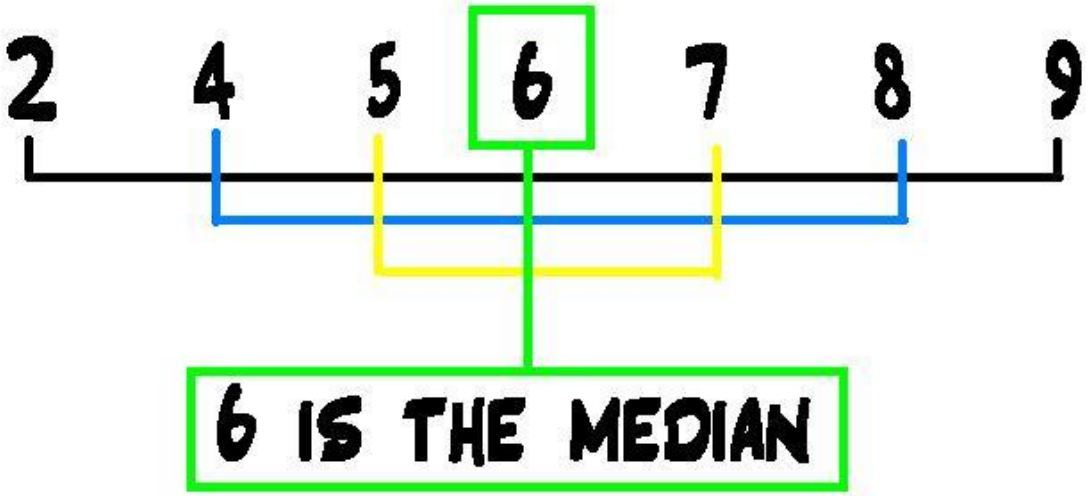
Calculation:

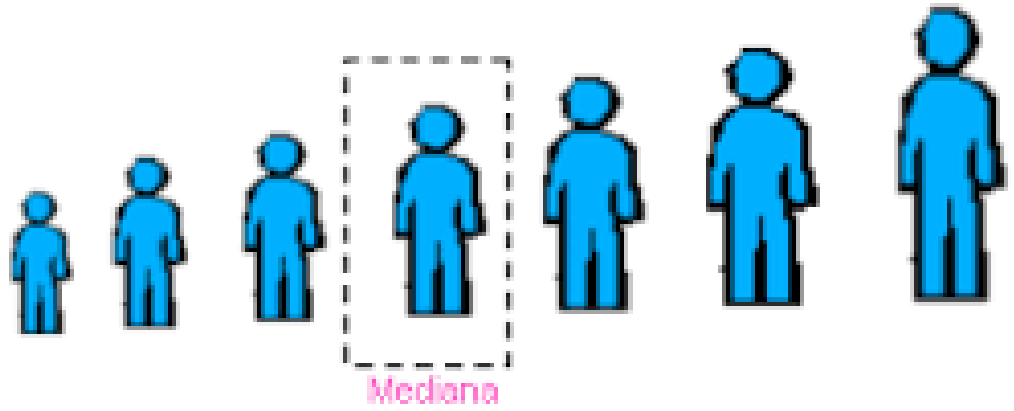
1. Arrange the observations from smallest to largest
2. Count in to find the middle value. The median is the middle value for an odd number of observations; it is defined as the mean of the two middle values for an even number of observations.

ODD number of observation

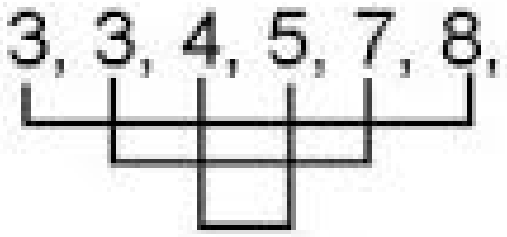
Given numbers: 8, 7, 4, 6, 5, 2, 9

Sort them than choose:





EVEN number of observation



4 + 5 = 9

9 / 2 =4.5

4.5 = Median

$$MEDIAN = \frac{D + I}{2}$$

Quartile/Percentile

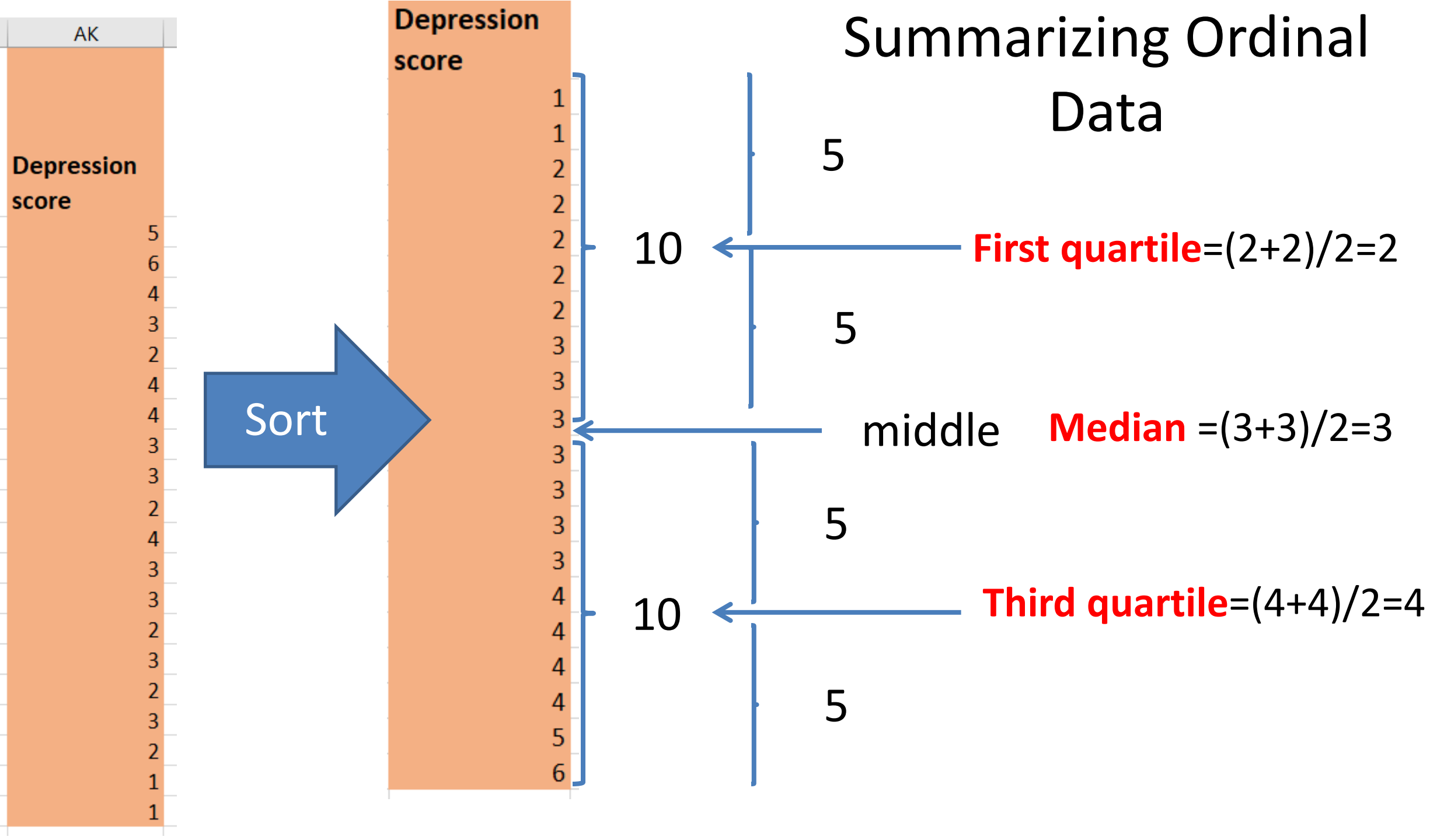
Divide the data in two parts:

- X% less than (<) Quartile/Percentile
- 100%-X% higher than (>) Quartile/Percentile

- Example: 25^{th} percentile=10
- 25% of data less than 10; 75% of data higher than 10

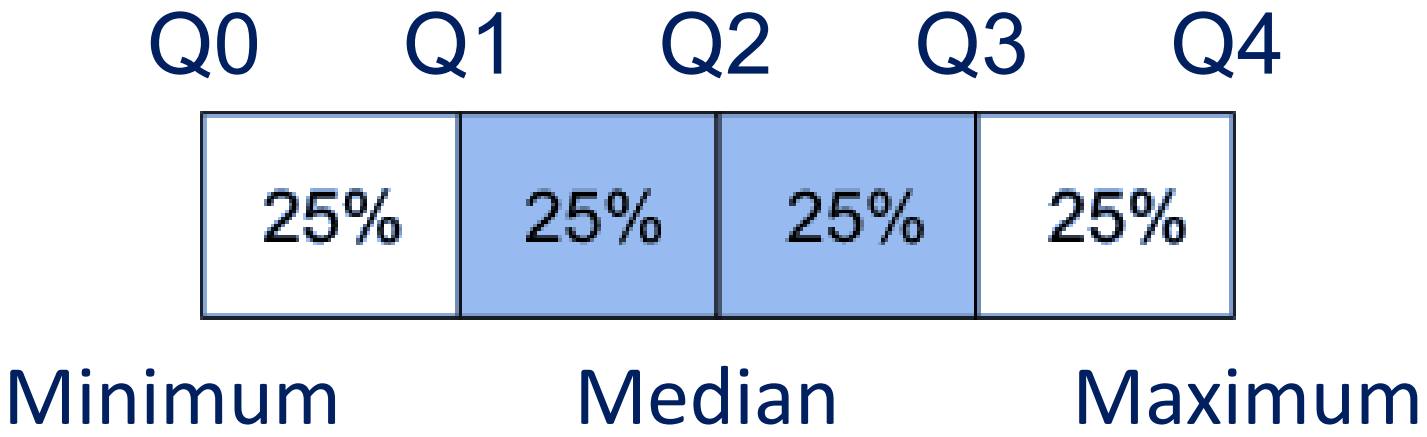
- 5 Quartiles
- 101 Percentiles

Summarizing Ordinal Data



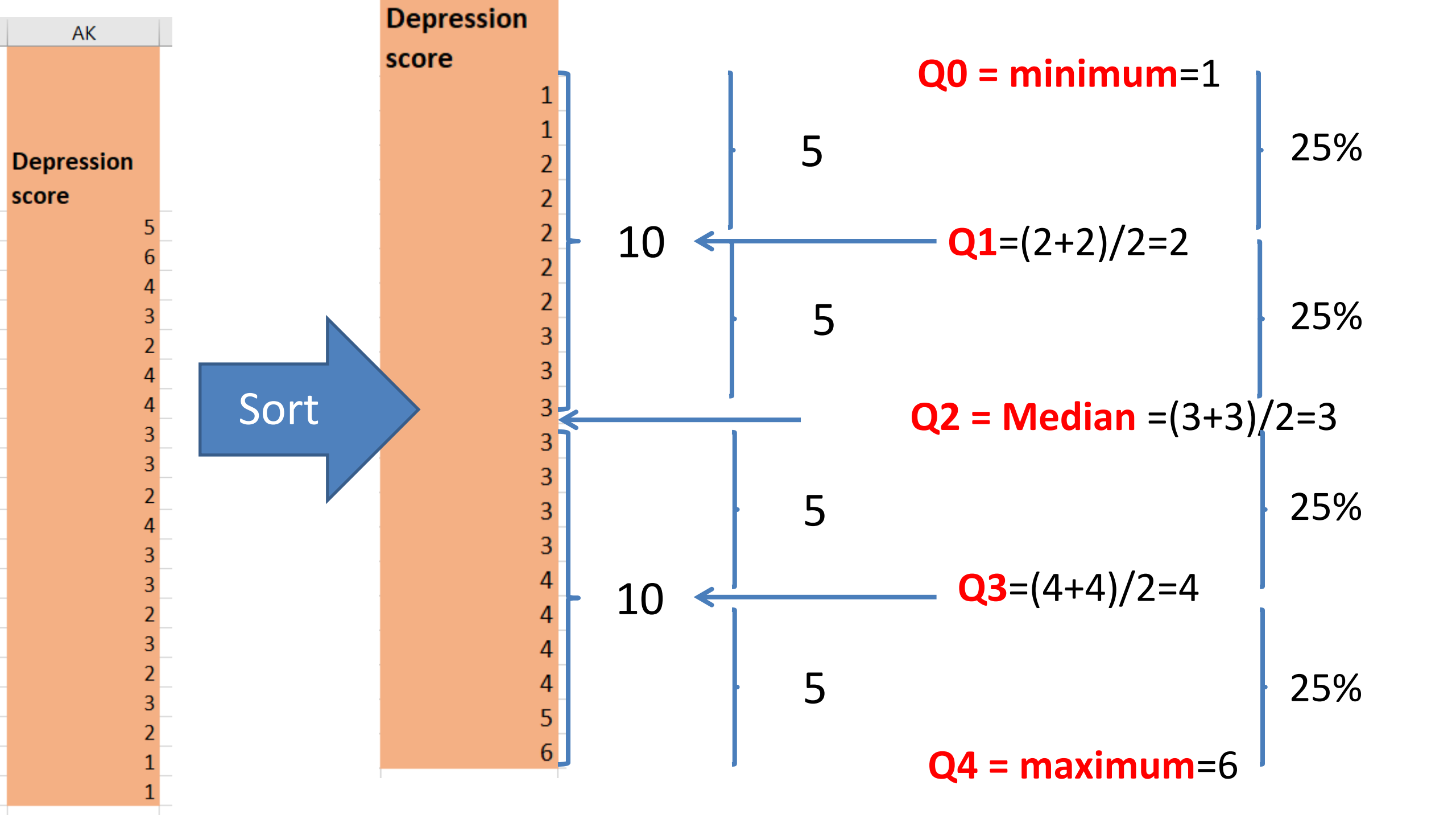
5 Quartiles

- Obs. Second quartile (Q2) = Median
- Q = quartile



5 Quartiles

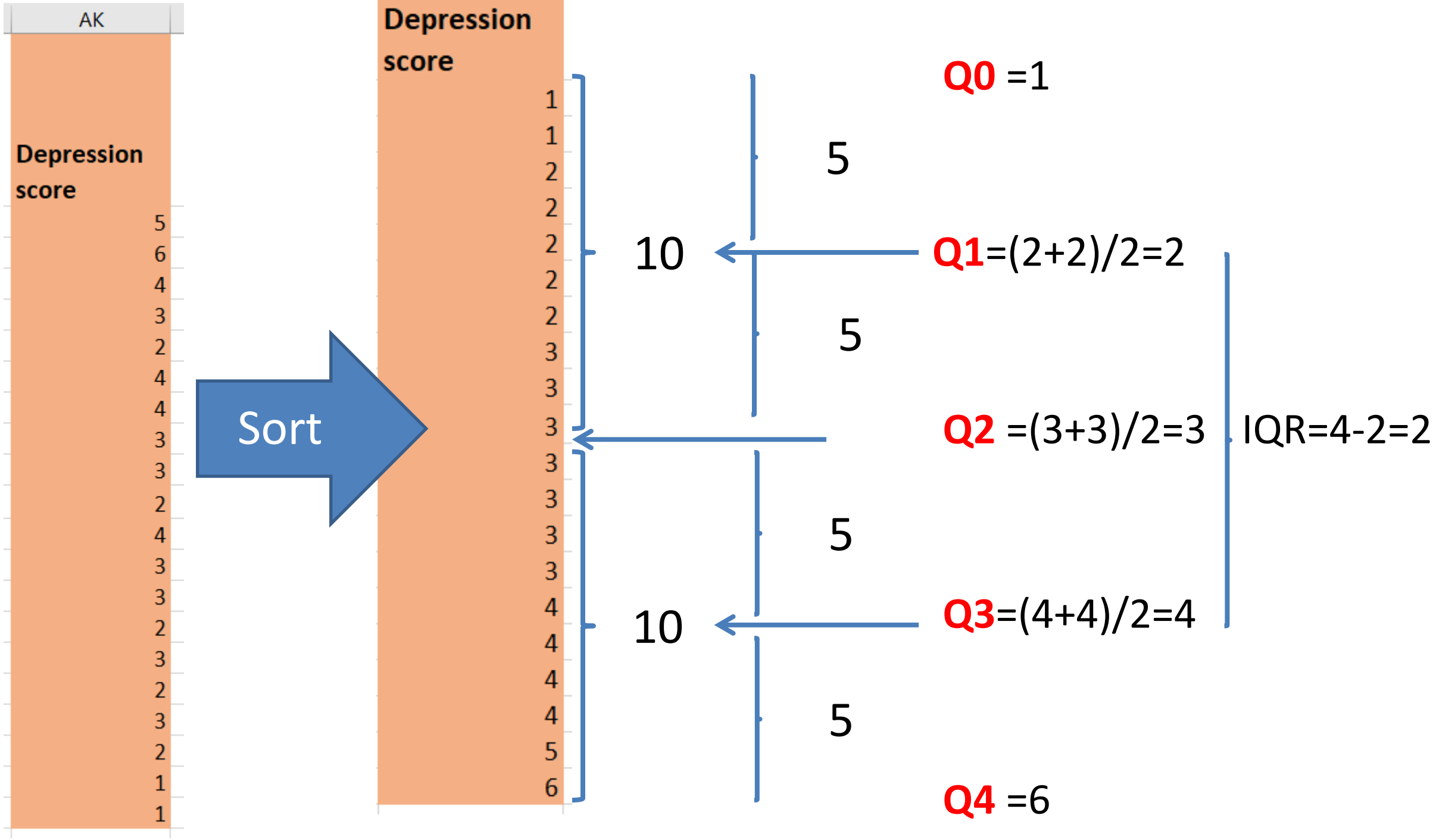
- **Quartile 0** = minimum
- **Quartile 1**
- **Quartile 2** = median
- **Quartile 3**
- **Quartile 4** = maximum


AK
Depression score
5
6
4
3
2
4
4
3
3
2
4
3
3
2
3
2
3
2
1
1
Sort
Depression score
1
1
2
2
2
2
2
3
3
3
3
3
3
3
3
4
4
4
4
5
6
Q0 = minimum=1
5
25%
10
Q1=(2+2)/2=2
5
25%
Q2 = Median =(3+3)/2=3
5
25%
10
Q3=(4+4)/2=4
5
25%
Q4 = maximum=6

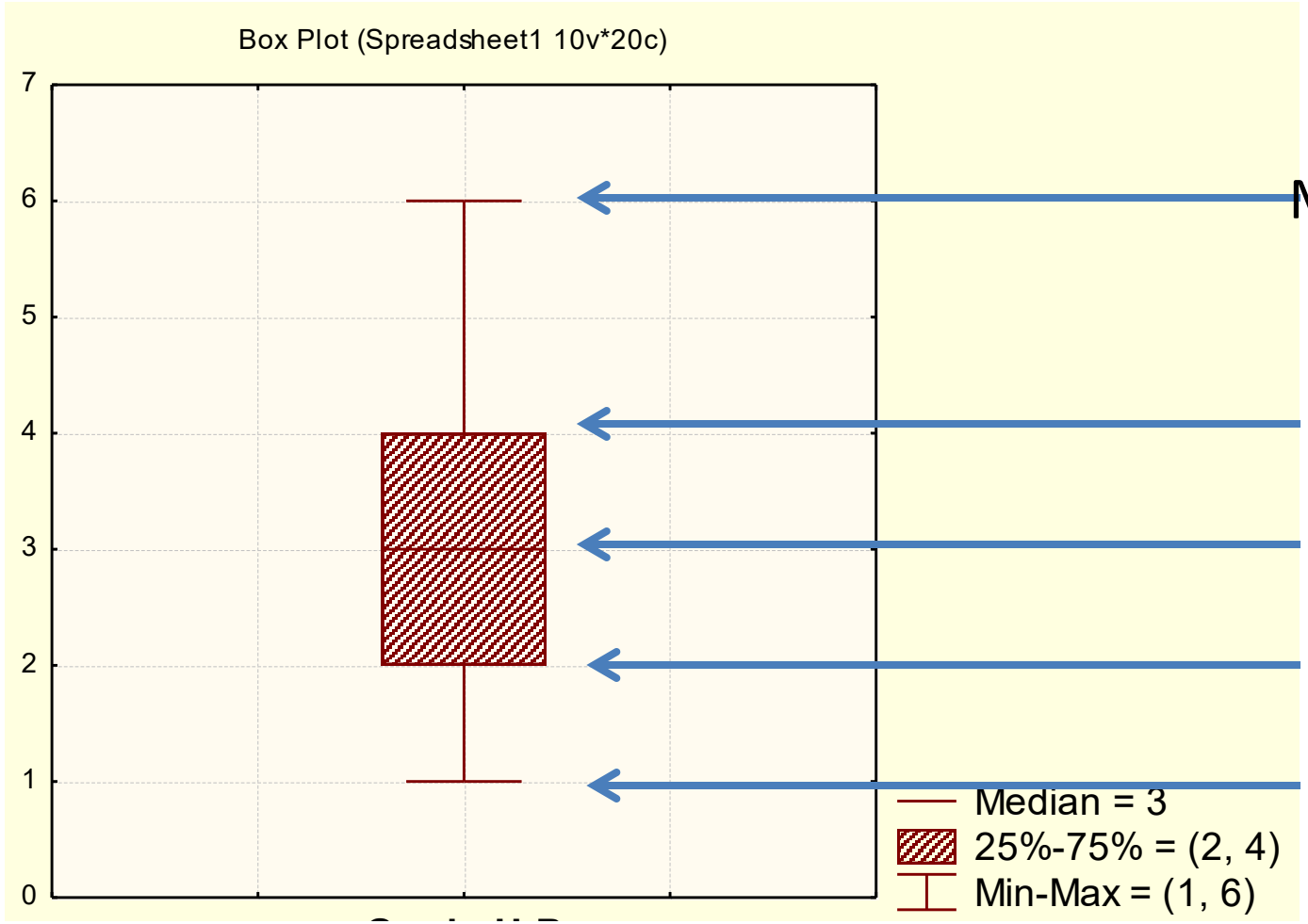
Other measures

- **Interquartile Range** - the difference between the third and first quartiles.

$$IQR = Q3-Q1$$


AK
Depression score
5
6
4
3
2
4
4
3
3
2
4
3
3
2
3
2
3
2
1
1
Sort
Depression score
1
1
2
2
2
2
2
3
3
3
3
3
3
3
3
4
4
4
4
5
6
Q0 =1
5
10
Q1=(2+2)/2=2
5
Q2 =(3+3)/2=3
IQR=4-2=2
5
10
Q3=(4+4)/2=4
5
Q4 =6

Ordinal Data – Chart – Box-whiskers Plot



Maximum = Quartile 4 = 6

25%

Quartile 3 = 4

Median = Quartile 2 = 3

25%

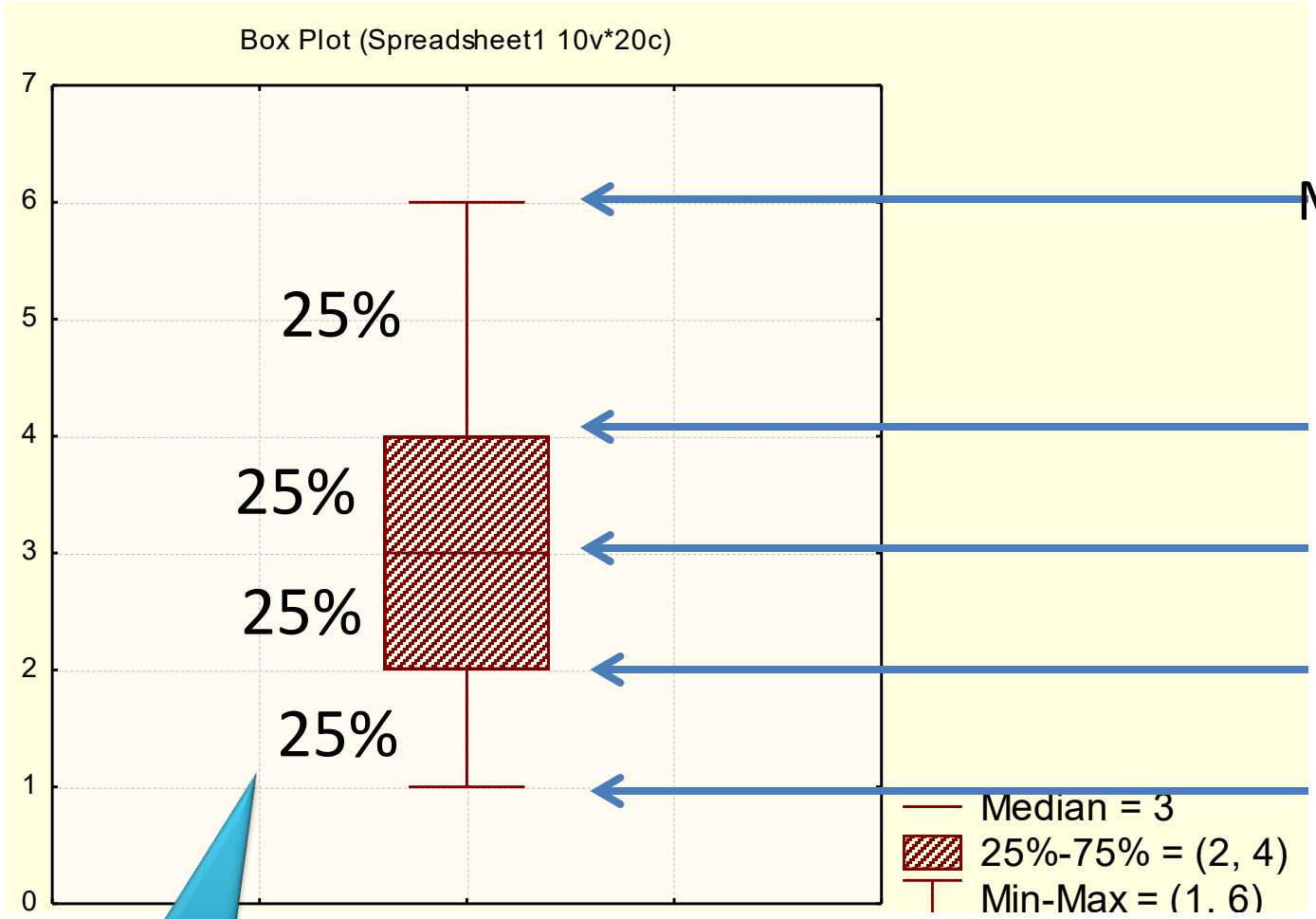
25%

Quartile 1 = 2

25%

Minimum = Quartile 0 = 1

Ordinal Data – Chart – Box-whiskers Plot



Maximum = Quartile 4 = 6

Quartile 3 = 4

Median = Quartile 2 = 3

Quartile 1 = 2

Minimum = Quartile 0 = 1

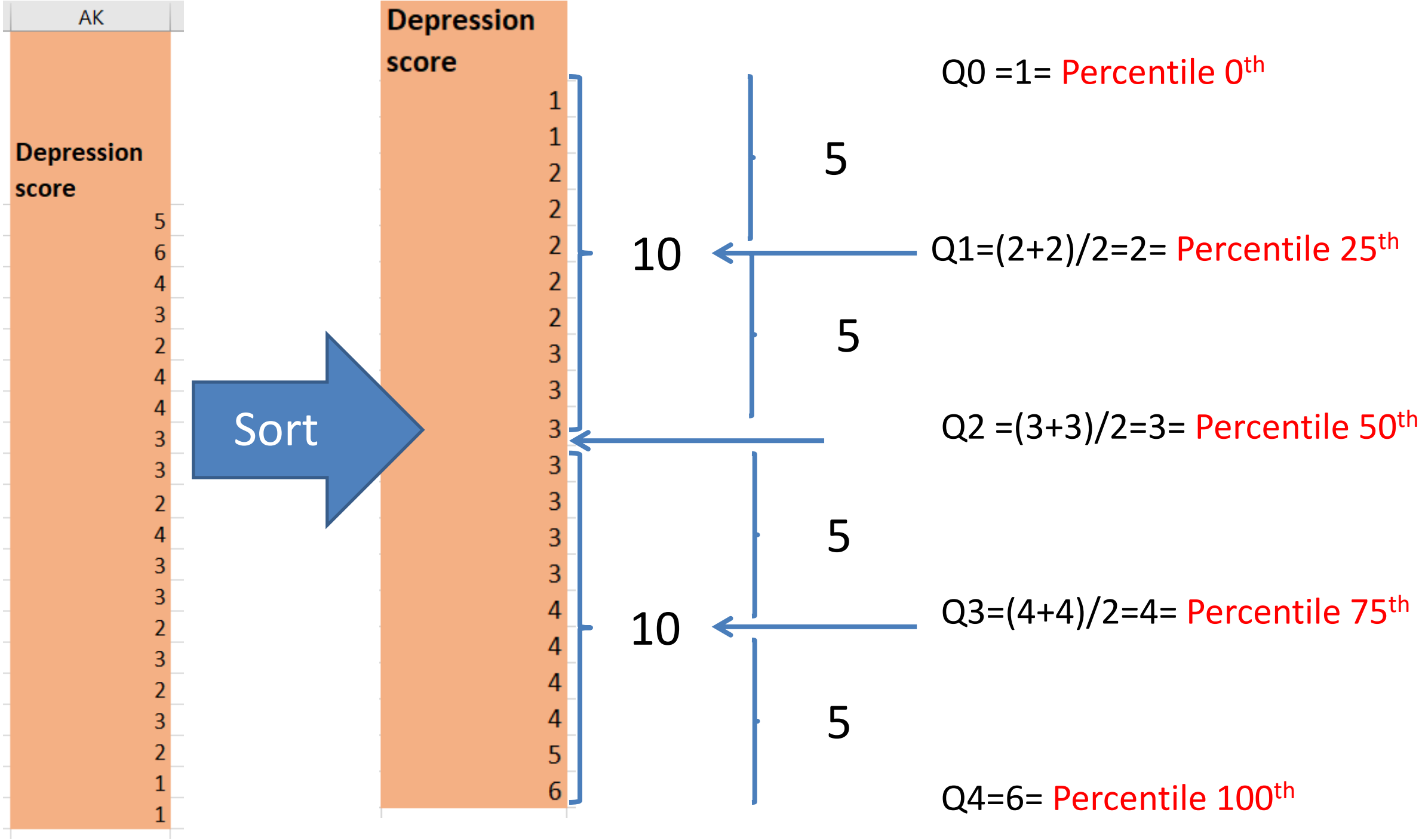
Same amount of data in each interval

Percentile

- the value below which, a given percentage of data in a group of data fall
- Ex. Percentile 25^{th} = 7,

  25% of data in the group are below 7

- 0-100 percentiles

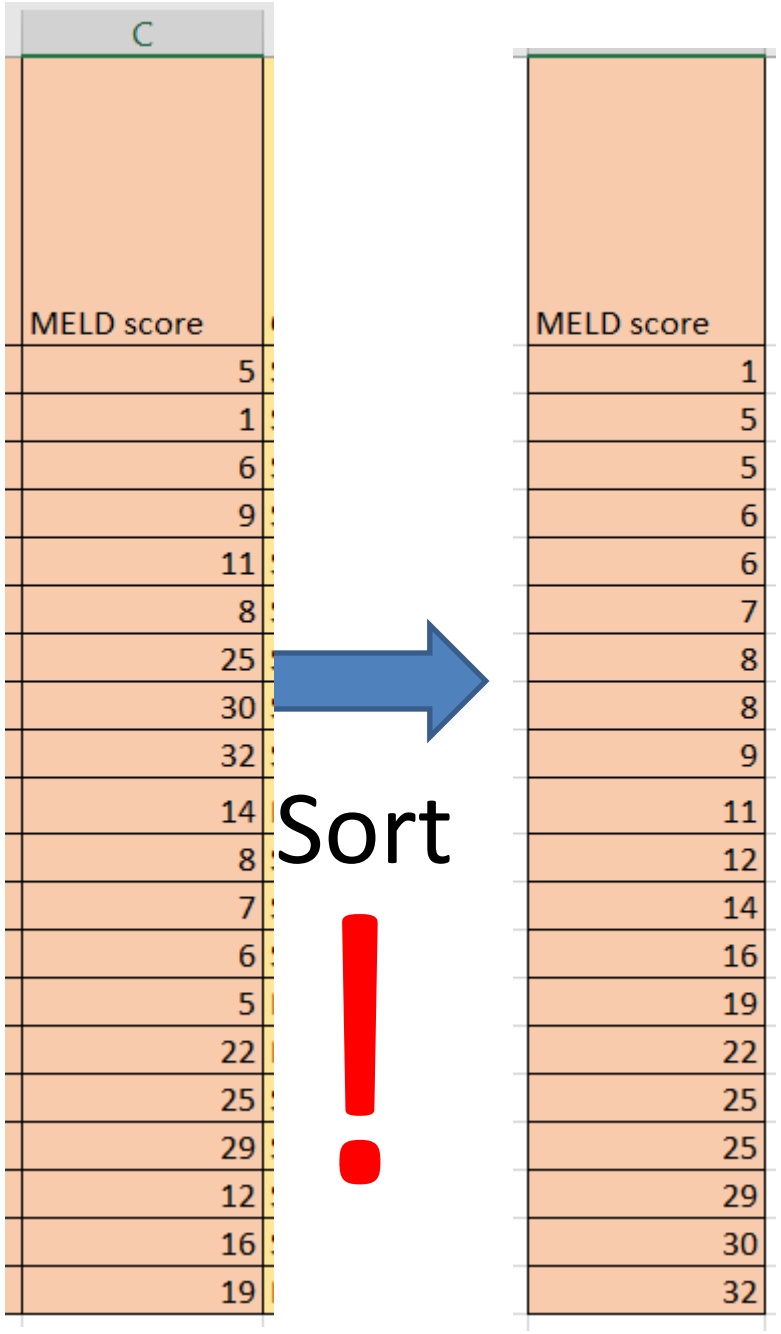

AK
Depression score
5
6
4
3
2
4
4
3
3
2
4
3
3
2
3
2
3
2
1
1
Sort
Depression score
1
1
2
2
2
2
2
2
3
3
3
3
3
3
3
3
4
4
4
4
5
6
5
10
5
10
5
Q0 =1= Percentile 0th
Q1=(2+2)/2=2= Percentile 25th
Q2 =(3+3)/2=3= Percentile 50th
Q3=(4+4)/2=4= Percentile 75th
Q4=6= Percentile 100th

Mode

- The most popular value! = Mode
- Marks for info exam for group 2: 10, 7, 5, 6, 10, 9, 8, 9, 10
- Most popular: 10 = mode

Summarizing Ordinal Data

- Parameters
  - Range
  - Median
  - Quartile 0-4
  - Interquartile range
  - Mode
- Chart
  - Box-Whiskers Plot
  - Column/Bar



Descriptive parameters

Minimum = 1

Maximum = 32

Range = 32 – 1 = 31

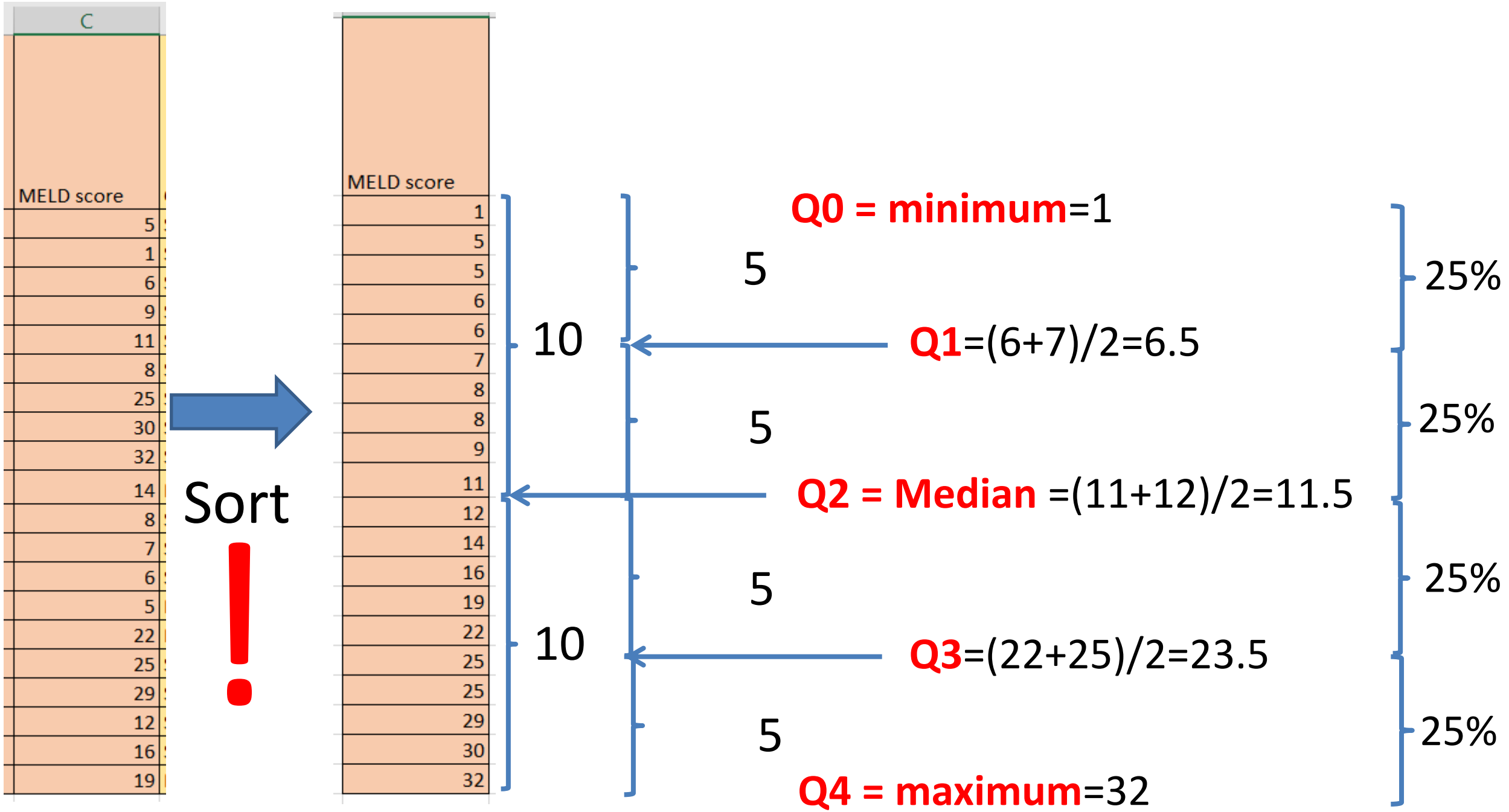
Median = (11 + 12)/2 = 11.5

Cuartile 1 = (6 + 7)/2 = 6.5

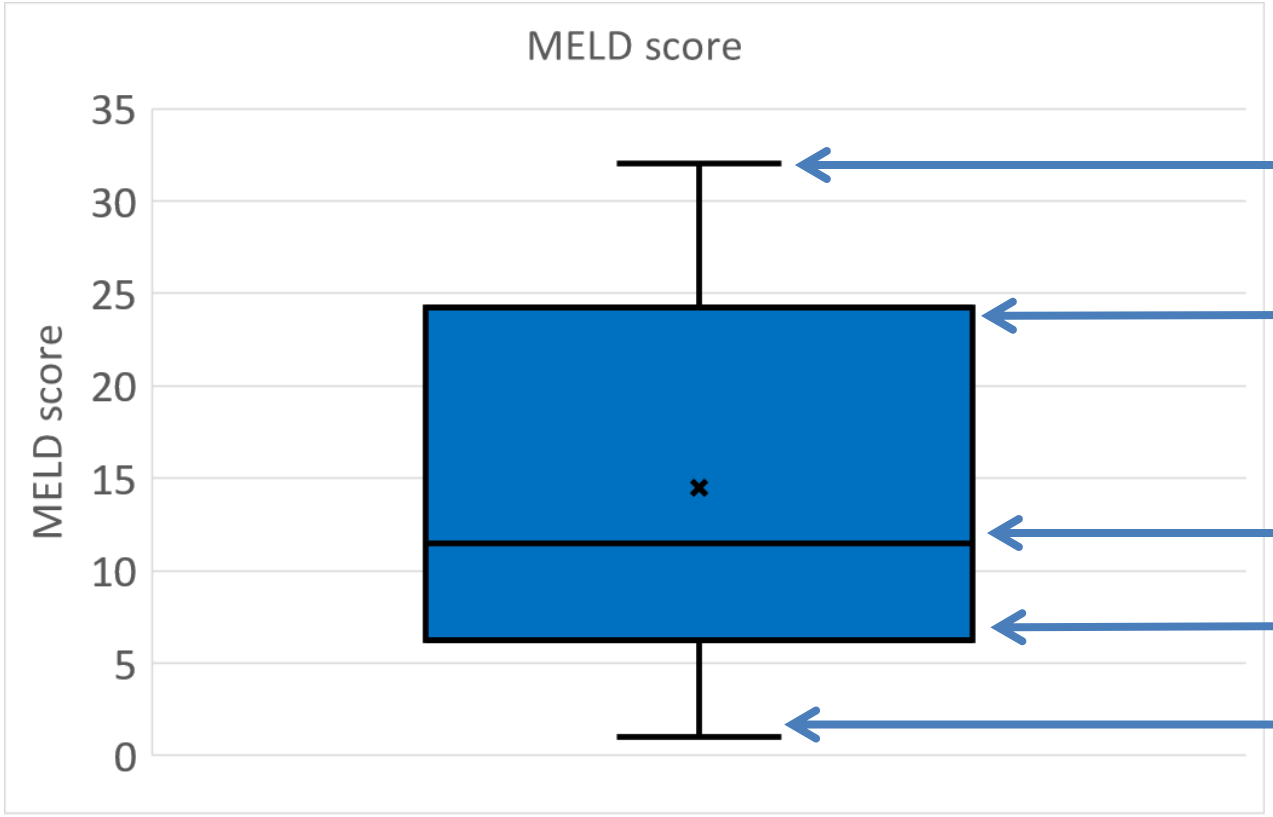
Cuartile 3 = (22 + 25)/2 = 47/2 =23.5

IQR = 23.5 – 6.5 = 17

Mode - none


C
MELD score
5
1
6
9
11
8
25
30
32
14
8
7
6
5
22
25
29
12
16
19
Sort
!
MELD score
1
5
5
6
6
7
8
8
9
11
12
14
16
19
22
25
25
29
30
32
10
10
Q0 = minimum=1
5
Q1=(6+7)/2=6.5
5
Q2 = Median =(11+12)/2=11.5
5
Q3=(22+25)/2=23.5
5
Q4 = maximum=32
25%
25%
25%
25%

Ordinal Data – Chart – Box-whiskers Plot



Maximum = Quartile 4 = 32

Quartile 3 = 23.5

Median = Quartile 2 = 11.5

Quartile 1 = 6.5

Minimum = Quartile 0 = 1

Relationship between a qualitative variable and a score (ordinal variable)

Parameters

B	C
Urban/Rural	MELD score
Urban	5
Rural	1
Urban	6
Urban	9
Urban	11
Rural	8
Urban	25
Urban	30
Rural	32
Rural	14
Urban	8
Urban	7
Urban	6
Rural	5
Rural	22
Urban	25
Urban	29
Urban	12
Rural	16
Rural	19

Sort by Urban/Rural than by MELD

!

Urban/Rural	MELD score
Rural	1
Rural	5
Rural	8
Rural	14
Rural	16
Rural	19
Rural	22
Rural	32
Urban	5
Urban	6
Urban	6
Urban	7
Urban	8
Urban	9
Urban	11
Urban	12
Urban	25
Urban	25
Urban	29
Urban	30

Rural subgroup

Urban subgroup

Urban/Rural	MELD score
Rural	1
Rural	5
Rural	8
Rural	14
Rural	16
Rural	19
Rural	22
Rural	32
Urban	5
Urban	6
Urban	6
Urban	7
Urban	8
Urban	9
Urban	11
Urban	12
Urban	25
Urban	25
Urban	29
Urban	30

Parameters

Urban group

number of people = 12

Minimum = 5

Maximum = 30

Range = 30 – 5 = 25

Median = (9 + 11)/2 = 10

25^{th} percentile = Q1 =(6 + 7)/2 = 6.5

75^{th} percentile = Q3 = (15 + 25)/2 = 25

IQR = 25 – 6.5 = 18.5

Mode = none

Rural group

number of people = 8

Minimum = 1

Maximum = 32

Range = 32 – 1 = 31

Median = (14 + 16)/2 = 15

25^{th} percentile = Q1 = (5 + 8)/2 = 6.5

75^{th} percentile = Q3 = (19 + 22)/2 = 20.5

IQR = 20.5 – 6.5 = 15.5

Mode - none

Urban/Rural	MELD score
Rural	1
Rural	5
Rural	8
Rural	14
Rural	16
Rural	19
Rural	22
Rural	32
Urban	5
Urban	6
Urban	6
Urban	7
Urban	8
Urban	9
Urban	11
Urban	12
Urban	25
Urban	25
Urban	29
Urban	30

Presenting the parameters in the papers

- Table. Descriptive statistics parameters for MELD score in rural versus urban group

	Minimum	Median (Q1 – Q2)	Maximum
Rural (n=8)	1	15 (6.5 – 20.5)	32
Urban (n=12)	5	10 (6.5 – 25)	30

Urban/Rural	MELD score
Rural	1
Rural	5
Rural	8
Rural	14
Rural	16
Rural	19
Rural	22
Rural	32
Urban	5
Urban	6
Urban	6
Urban	7
Urban	8
Urban	9
Urban	11
Urban	12
Urban	25
Urban	25
Urban	29
Urban	30

Chart – Box-whiskers plot

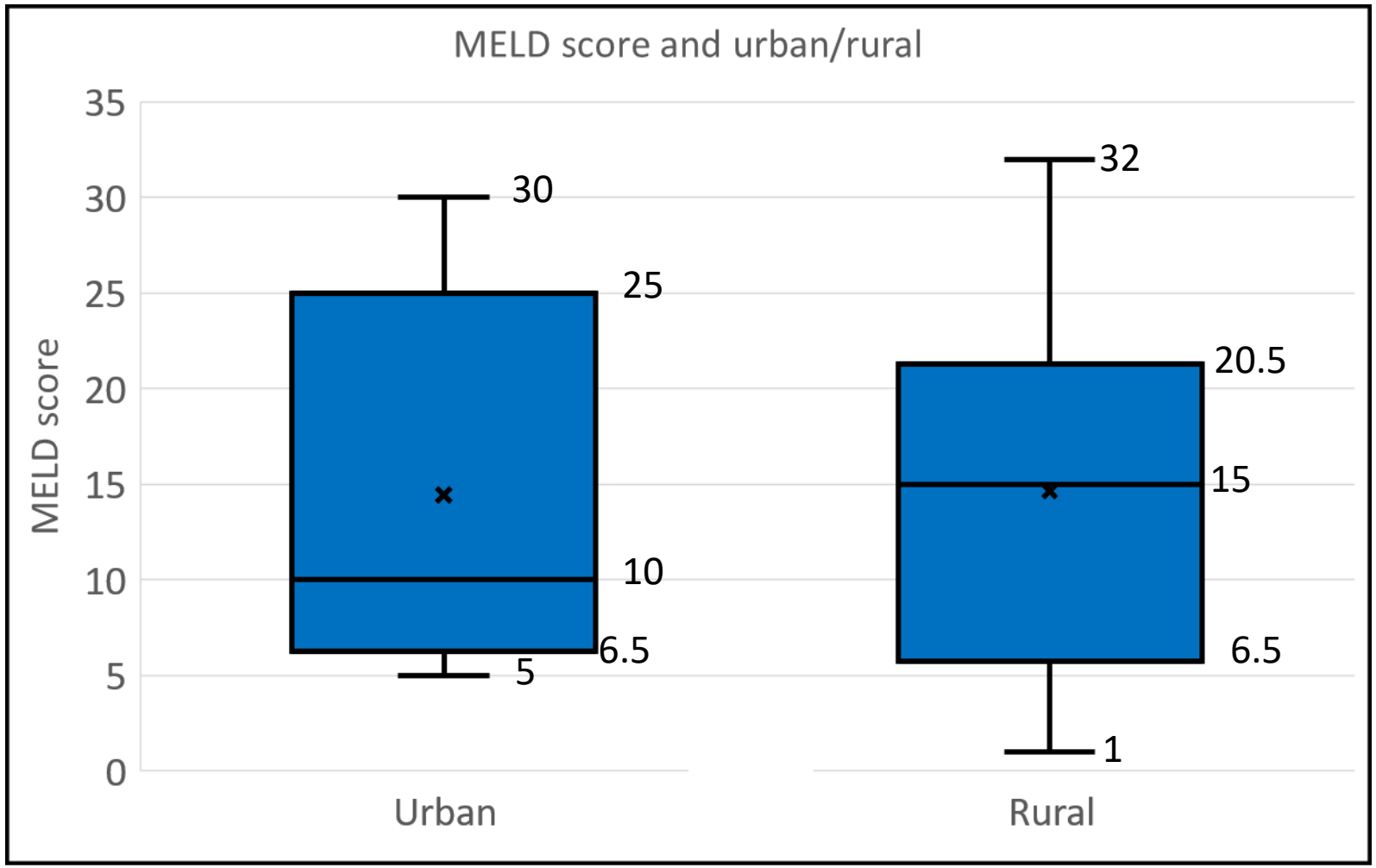


Figure. MELD score distribution in urban versus rural

Calculate the parameters in Excel

range C2:C21

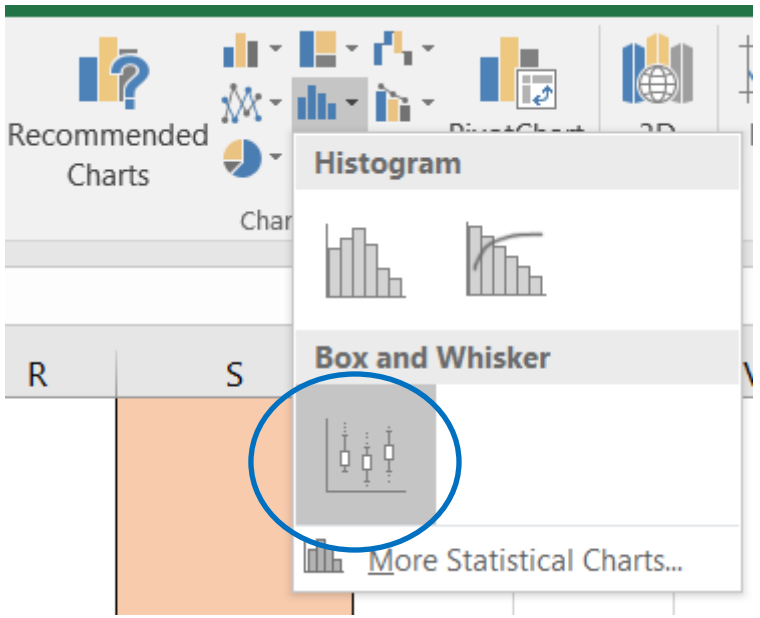
	C
1	MELD score
2	5
3	1
4	6
5	9
6	11
7	8
8	25
9	30
10	32
11	14
12	8
13	7
14	6
15	5
16	22
17	25
18	29
19	12
20	16
21	19

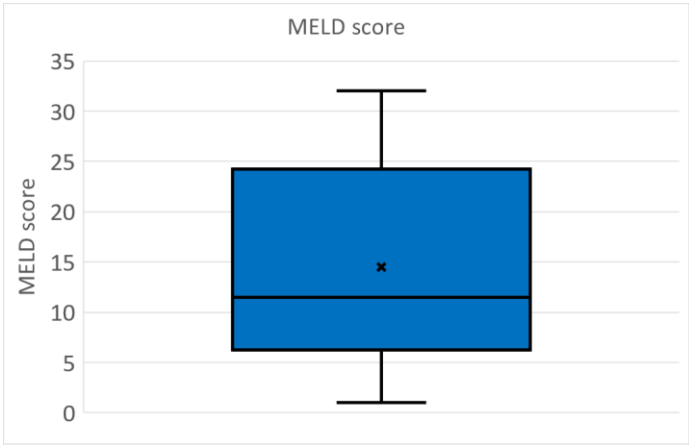
Parameter	Equation in Excel
Minimum	=min(C2:C21)
Maximum	=max(C2:C21)
Range	=max(C2:C21)-min(C2:C21)
Median	=median(C2:C21)
Cuartile 1	=quartile(C2:C21,1)
Cuartile 3	=quartile(C2:C21,3)
Mode	=mode(C2:C21)

Box-plot in Excel for one ordinal variable

Select the data

	C
1	MELD score
2	5
3	1
4	6
5	9
6	11
7	8
8	25
9	30
10	32
11	14
12	8
13	7
14	6
15	5
16	22
17	25
18	29
19	12
20	16
21	19

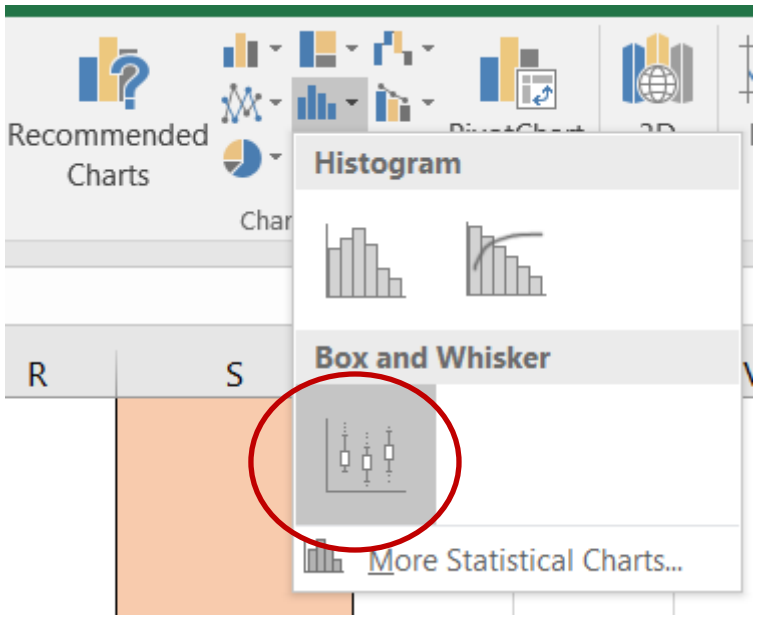


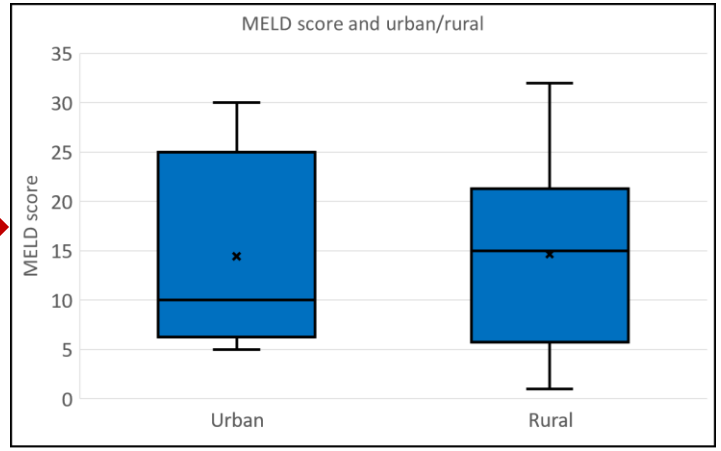


Box-plot in Excel for one ordinal variable

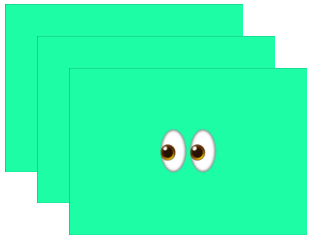
Select both columns

B	C
Urban/Rural	MELD score
Urban	5
Rural	1
Urban	6
Urban	9
Urban	11
Rural	8
Urban	25
Urban	30
Rural	32
Rural	14
Urban	8
Urban	7
Urban	6
Rural	5
Rural	22
Urban	25
Urban	29
Urban	12
Rural	16





Summarizing Ordinal Variables



https://www.wooclap.com/XZLYSC

One variable

- Parameters
  - Range
  - Median
  - Quartile 0-4
  - Interquartile range
  - Mode
- Chart
  - Box-Whiskers Plot
  - Column

One dichotomial variable and one ordinal variable

- Parameters on each subgroup
  - Range
  - Median
  - Quartile 0-4
  - Interquartile range
  - Mode
- Chart
  - Box-Whiskers Plot
  - Column/Bar
  - Staked Column/Bar
  - 100% Stacked Column/Bar

Key concepts

- **Frequency tables** display the number of observations with a specific qualitative characteristic
- A **contingency table** describes the relationship between two qualitative characteristics
- **Pie charts** represent percentages (relative frequencies)
- **Column/bar charts** represent absolute frequencies
- **Stacked column/bar charts** depict relationships between two qualitative characteristics

Examples from literature

Medical scientific articles example

Example A

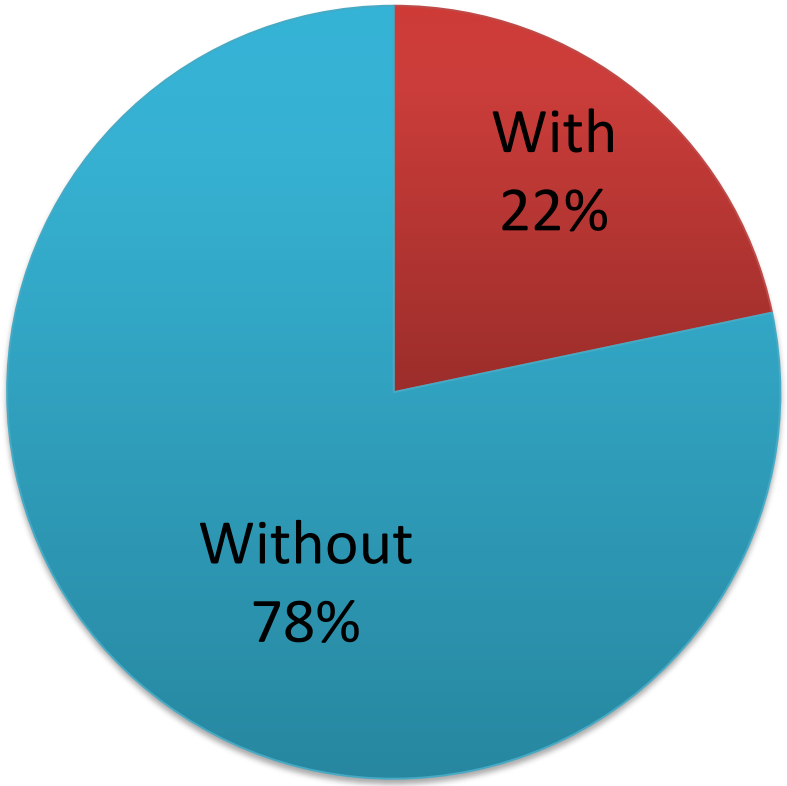
- **Background:**
- needle fear,
- needle anxiety
- needle phobia

- **Aim:** The study want to answer to the question:
  - "How many young people are afraid of injection?"

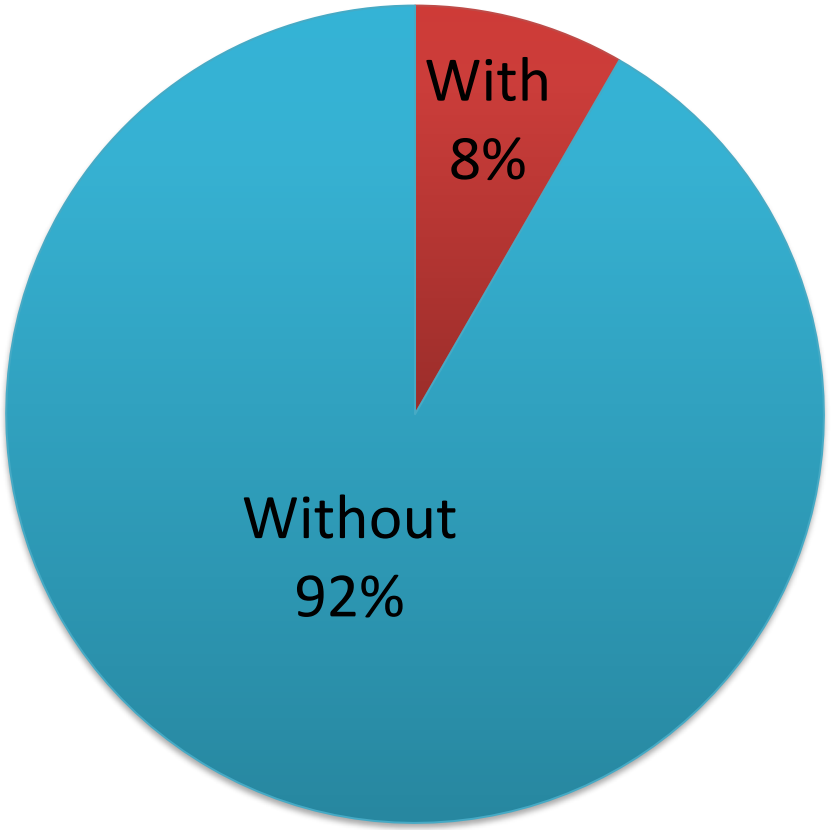
- **Materials and Methods:** 400 travelers visiting a travel health clinic were evaluated

Nir Y, Paz A, Sabo E, Potasman I. Fear of injections in young adults: prevalence and associations. Am J Trop Med Hyg. 2003 Mar;68(3):341-4.

Injection fear - 21.7%



Intense injection fear – 8.2%

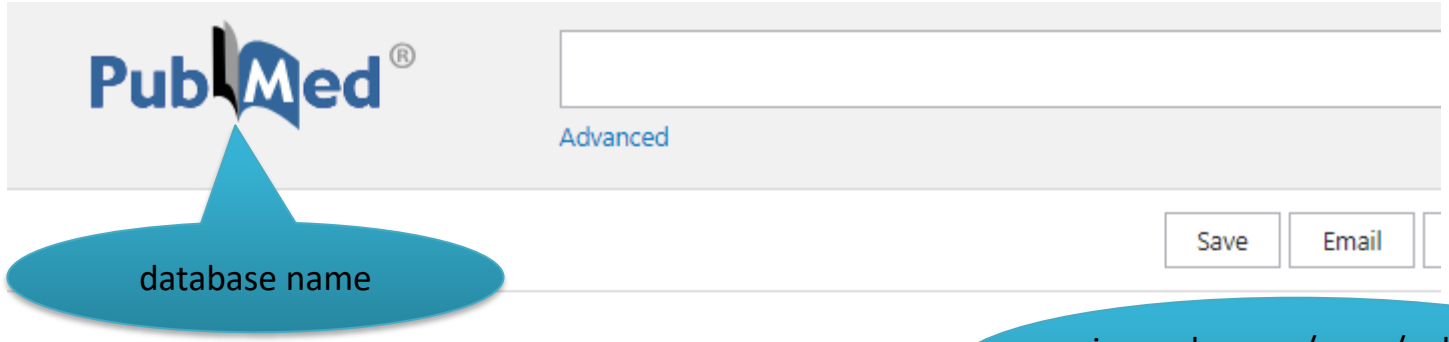


Nir Y, Paz A, Sabo E, Potasman I. Fear of injections in young adults: prevalence and associations. Am J Trop Med Hyg. 2003 Mar;68(3):341-4.

Causes of injection fear

- Factors
- watching other people being vaccinated,
- fear of pain,
- needle size,
- a history of fainting

Nir Y, Paz A, Sabo E, Potasman I. Fear of injections in young adults: prevalence and associations. Am J Trop Med Hyg. 2003 Mar;68(3):341-4.



journal name /year /volum number/ issue

Am J Trop Med Hyg. 2003 Mar;68(3):341-4.

Fear of injections in young adults: prevalence and associations

Title of the articles

Yael Nir [1], Alona Paz, Edmond Sabo, Israel Potasman

Affiliations + expand

PMID: 12685642

authors

Abstract

Fear of injections may interfere with receipt of vaccines. The frequency, associations, and precipitators of fear-provoking factors of 400 travelers visiting a travel health clinic were evaluated. The median age of this group was 25, 7% were medical staff members, and 2.8% were regular injectors (insulin). Eighty-five (21.7%; 95% confidence interval, 17.3-25.6%) of the travelers indicated that they were afraid of injections, and in 8.2%, the fear was unreasonably intense. Multivariate analysis revealed that watching other people being vaccinated, fear of pain, needle size, and a history of fainting were highly and independently associated with injection phobia. The sensitivity, specificity, and discrimination accuracy of this model were 79.5%, 78.0%, and 78.3%, respectively. Injection phobia and a bad past vaccination experience were significantly associated with fainting. Perceived empathy, on the other hand, was a significant protective factor. Fear of injections was common in this cohort and was highly associated with past fainting after vaccination.

abstract=article summary

Journal of Indian Society of Periodontology

Home
Current issue
Instructions
Submit article

J Indian Soc Periodontol. 2014 May-Jun; 18(3): 369–374. PMCID: PMC4095633
doi: 10.4103/0972-124X.134580 PMID: 25024553

where and when is publish: journal name, volum number/ issue, year

Evaluating awareness regarding oral hygiene practices and exploring gender differences among patients attending for oral prophylaxis

Title of the articles

Sukhvinder Singh Oberoi, Vikrant Mohanty,[1] Ananya Mahajan,[1] and Avneet Oberoi[2]

authors

Abstract

Go to:

Background:

Oral hygiene is intimated in health of all parts of the body including oral cavity. The understanding of actual practices in keeping the oral heath at standard based on patient's perceptions of oral health care is vital. Understanding the effect of gender on oral health would facilitate the development of successful attitude and behavior modification approach towards sustainable oral health.

Purpose of Study:

To evaluate awareness regarding oral hygiene practices and exploring gender differences among patients attending for oral prophylaxis.

Materials and Methods:

A survey was conducted among 250 patients attending the department of periodontology, Maulana Azad institute of dental sciences for oral prophylaxis. A structured questionnaire was used to collect information regarding practices and perception about oral hygiene.

Results:

Majority of the patients (60.4%) felt that oral hygiene is mandatory for overall health of the body. The use of toothpaste and toothbrush (83.6%) was the most preferred cleaning aid among the study population in the present study. The major constraint for avoiding dental examination was no felt need (41.2%) followed by cost of dental treatment (26.8%) and time constraints (24.0%).

Conclusions:

Professional plaque removal and regular follow-up combined with oral hygiene instructions to the patients can minimize the level of gingival inflammation and swelling. The poor resources for dental care, common malpractices and nonavailability of professional care are the main barriers in seeking optimum oral hygiene.

abstract=article summary

Table 2. Distribution of the oral hygiene practices according to gender

Questions	Response	Male (%)	Female (%)	Total (%)
Question 1: Do you think oral hygiene is mandatory for overall health of the body?	Yes	86 (64.1)	65 (56.0)	151 (60.4)
	No	8 (6.0)	6 (5.2)	14 (5.6)
	Don't know	40 (29.9)	45 (38.8)	85 (34.0)

Table title

what is in () it will be in %= relative frequency

Question 1 = a variable

Possible answers = cathegories of the variable

absolute frequency

relative frequency

Prevalence of burnout among German radiologists: a call to action

Radiological Education | Open access | Published: 12 February 2024
Volume 34, pages 5588–5594, (2024) Cite this article

Download PDF

You have full access to this open access article

Title

Moritz B. Bastian, Laureen Fröhlich, Joel Wessendorf, Michael Scheschenja, Alexander M. König, Jarmila Jedelska & Andreas H. Mahnken

Authors

Abstract

Objectives

In the presence of escalating global concerns regarding physician burnout, this study aims to analyze the prevalence and associated factors of burnout among radiologists in Germany.

Methods

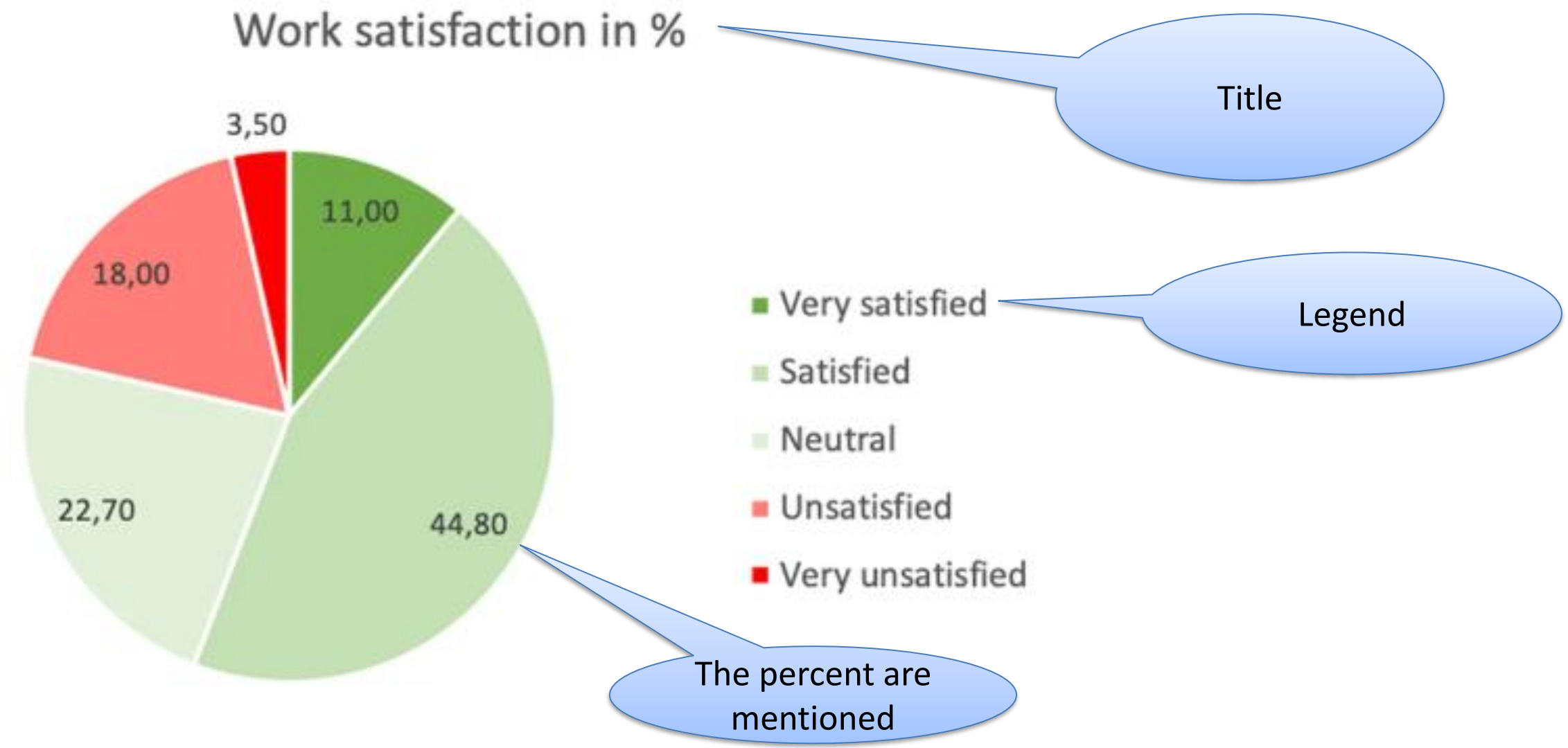
A comprehensive online survey, inclusive of 73 targeted questions including a German-modified version of the Maslach Burnout Inventory, was distributed among all members of the German Radiological Society and the German Interventional Radiological Society between May and August 2023. The survey encompassed aspects of employment, workload, well-being, and coping mechanisms. Data from 172 completed surveys were analyzed, with correlations explored via crosstabs and the Pearson-chi-square test.

Results

In total, 76.7% of participating radiologists were identified to be burnt out. The prevalence was significantly associated with increased workload, reduced sleep quality, suboptimal working conditions, reduced job satisfaction, and the negative interplay between work, family life, and health. Median work satisfaction was described as "satisfied" while median workload was assessed as "frequently overwhelming of work." A total of 41.9% of respondents noted facing daily time pressure. Radiologists' concerns about work interfering with private family life were voiced by approximately 70%, and 73.3% highlighted the perceived negative effects on their health.

abstract=article summary

Bastian, M.B., Fröhlich, L., Wessendorf, J. *et al.* Prevalence of burnout among German radiologists: a call to action. *Eur Radiol* **34**, 5588–5594 (2024).



- The questionnaire answers of the 172 radiologist : “How satisfied are you with your work?”

Mortality in USA

Centers for Disease Control and Prevention
CDC 24/7: Saving Lives, Protecting People™

National Center for Health Statistics

CDC > NCHS > Publications and Information Products > Data Briefs

Publications and Information Products

Data Briefs

Death in the United States, 2010

Health E-Stats

Health, United States

Life Tables

National Health Statistics Reports

National Vital Statistics Reports

Vital and Health Statistics Series

Death in the United States, 2010

NCHS Data Brief No. 99, July 2012

PDF Version (833 KB)

Arialdi M. Miniño, M.P.H., and Sherry L. Murphy, B.S.

Key findings

Data from the National Vital Statistics System (Mortality)

- Life expectancy at birth is 78.7 years. Hispanic females have the longest life expectancy (83.8 years) followed by non-Hispanic white females (81.1 years).
- The largest decrease in mortality between the years 2000 and 2010 occurred in the age group under age 25 years (15.8 percent), followed by those aged 65 years and over (13.3 percent).

- Minino Am, Murphy SL. Death in United States, 2010. NCHS Data Brief. 2012:99.

Accidents, for example, accounted for 38 percent of all deaths among persons aged 1–24 years. Heart disease, on the other hand, accounted for only 3 percent of all deaths in this same age group. For persons aged 65 years and over, heart disease ...ause of death, accounting for 27 percent of all deaths in this age group, whereas accidents are not among the ...s of death for those aged 65 years and over (Figure 4).

Figure has a title

Figure is mentioned in the text

Figure 4. Percent distribution of five leading causes of death, by age group: United States, preliminary 2010

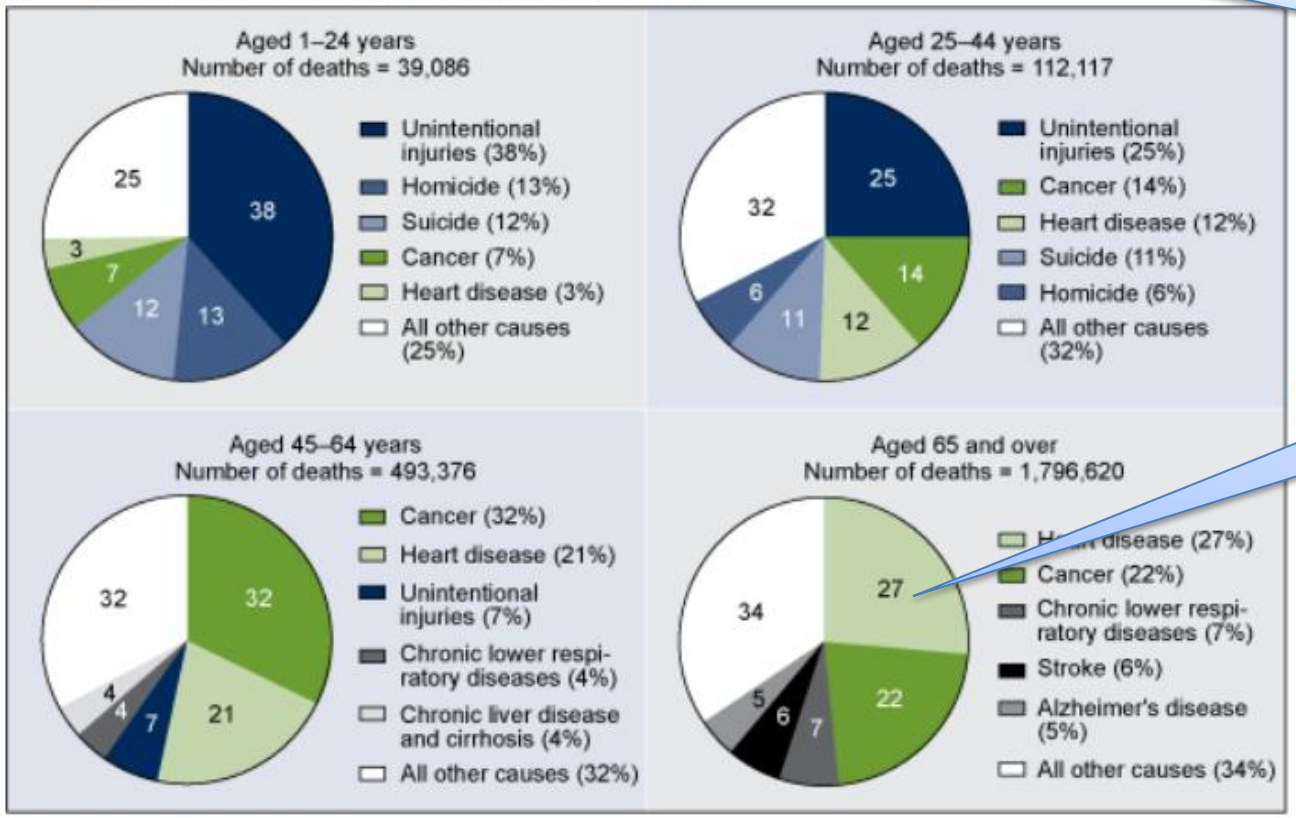


Figure has a brief description

Percentages are measured

SOURCE: National Vital Statistics System, Mortality.

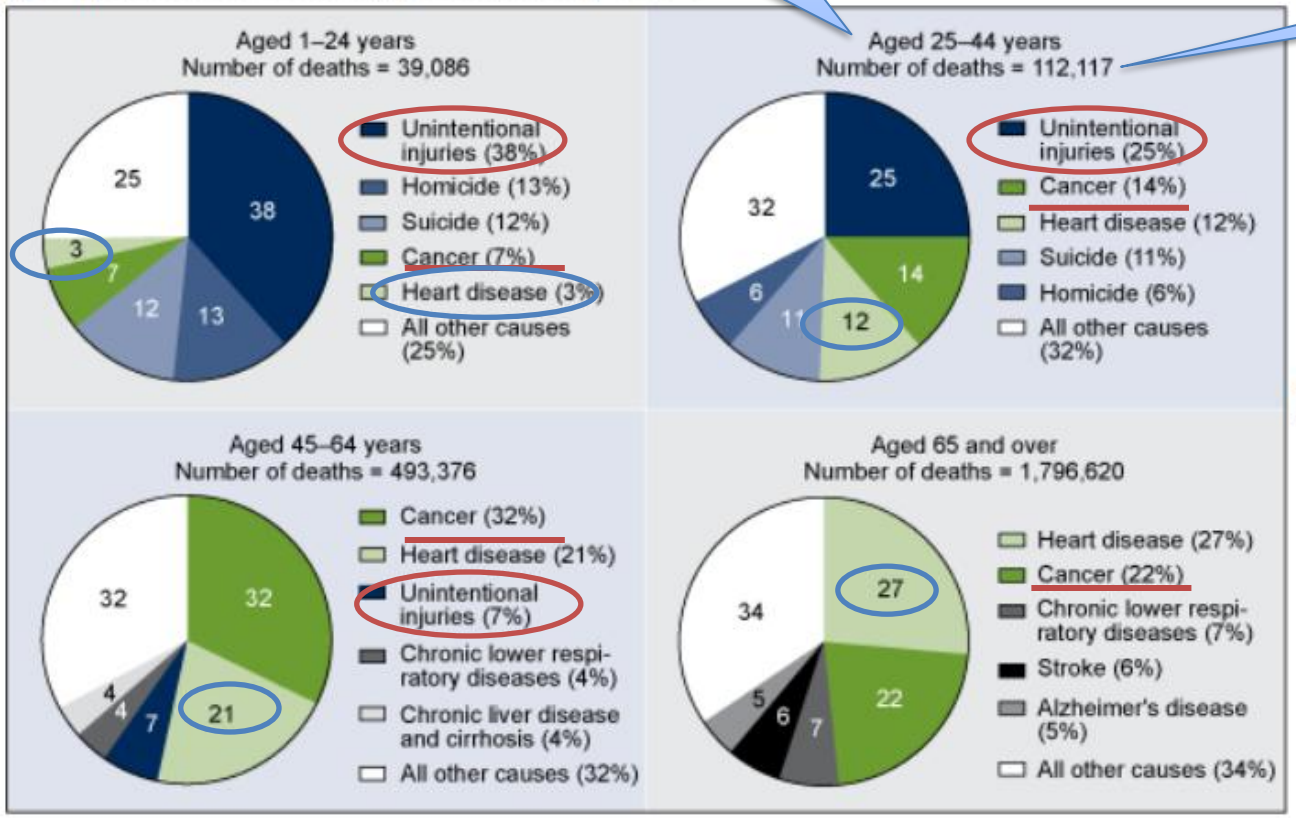
Minino Am, Murphy SL. Death in United States, 2010. NCHS Data Brief. 2012:99.

Accidents, for example, accounted for 38 percent of all deaths among persons aged 1–24 years. Heart disease, on the other hand, accounted for only 3 percent of all deaths in this same age group. For persons aged 65 years and over, heart disease is th[illegible]rcent of all deaths in this age group, whereas accidents are not among the [illegible]ver (Figure 4).

Age category = ordinal qualitative variable with 4 categories

Number of deaths increase with age

Figure 4. Percent distribution of fiv[illegible]g causes of death, by age group: United States, preliminary 2010



Interpretations: At a younger age the most common cause of death ib by unintentional injuries

At the age between 45 to 65 the most probable cause of death is from cancer

The risk of dying from cardiovasculare disease increase with age

SOURCE: National Vital Statistics System, Mortality.

Minino Am, Murphy SL. Death in United States, 2010. NCHS Data Brief. 2012:99.

Optional reading

- **Rate** = a / n*base, eg. Base 1000, 10000, 100000
  The mortality rate = number of deaths in a population / no. individuals in the population
- **Fatality rate =** number of deaths of people with a disease / number of sick people
- **Infant mortality rate** = number of infants who died up to one year old / 1,000 births
  Morbidity rate = an inexact term that can mean either the incidence rate or the prevalence rate.
  Prevalence - no. of sick people at a time / Total number people at risk
  Incidence - no. of newly diagnosed cases in a given period / population at risk at the beginning of interval

Optional Exercices 1

D
Maxilary
superior
inferior
superior
inferior
superior
inferior
superior
inferior
inferior
superior
superior
inferior
inferior
inferior
inferior
inferior
inferior
superior
inferior
superior
superior

- What kind of chart can be used?
- Make the frequency table.

Response

- If you want to know which chart to choose – you should know the type of variable
- Pie chart is preferable
- It is not recommended a Pie chart with more than 6 slices.

Theoretical exam – example of questions

- Which of the following are an ordinal variable?
- A. teeth movement (yes/no)
- B. difficulty of the procedure
- C. stage of the cancer
- D. fluoride intake per day in grams
- E. number of teeth brushings per day

Theoretical exam – example of questions

- Which of the following are an ordinal variable?
- A. teeth movement (yes/no)
- B. **difficulty of the procedure**
- C. **stage of the cancer**
- D. fluoride intake per day in grams
- E. number of teeth brushings per day

- *Let be a statistical series with the following data: 40, 60, 20, 20, 60, 80, 80, 40, 60 and 90. The relative frequency of 0.3 corresponds to:
- A. 20
- B. 80
- C. 40
- D. 60
- E. none is correct

- *Let be a statistical series with the following data: 40, 60, 20, 20, 60, 80, 80, 40, 60 and 90. The relative frequency of 0.3 corresponds to:
- A. 20
- B. 80
- C. 40
- **D. 60**
- E. none is correct

- In a study the researcher observe the color of teeth on a scale with 8 different colors. They repeat the experiment for 300 patients. They obtain 300 different values, one for each patient. Which kind of chart is suitable to represent the distribution of color frequencies in this patients?
- A. bar of pie chart
- B. bar chart
- C. histogram
- D. pie chart
- E. scatter XY chart

- In a study the researcher observe the color of teeth on a scale with 8 different colors. They repeat the experiment for 300 patients. They obtain 300 different values, one for each patient. Which kind of chart is suitable to represent the distribution of color frequencies in this patients?
- A. bar of pie chart
- B. **bar chart**
- C. histogram
- D. **pie chart**
- E. scatter XY chart

Thank you!