
DESCRIPTIVE STATISTICS I

Sorana D. Bolboacă

OBJECTIVES

- Representation of qualitative data
- Absolute and relative frequencies
- Frequency tables
- Graphical representations

DESCRIPTIVE VS INFERENCEAL STATISTICS

- Descriptive statistics:
 - Summarize or describe a collection of data
- Inferential statistics:
 - Used to draw inferences about a population from a sample:
 - Estimation: parameter and confidence interval
 - Hypothesis testing (null and alternative hypotheses): determine whether the data are strong enough to reject the null hypothesis

SUMMARIZING MEDICAL DATA

- Large amounts of medical data are compressed into more easily assimilated summaries
 - Provide the user with a sense of the content
- There a number of ways data can be presented depending by the type of variables:
 - Qualitative variables
 - Quantitative variables

GOOD TABLES PRACTICES

1. Simple: it is preferred to have 2 or 3 small tables instead of one big table
2. Must be information without reading the associated text:
 - ❑ Abbreviations and symbols must be explained at the bottom of the table
 - ❑ Definitions of rows and columns with units of measurements in headings (if it is applied)
 - ❑ Brief descriptive heading: what? when? where?
 - ❑ Must not duplicate material in the text or in illustration
 - ❑ Synthesis (total) rows and columns
3. If data are taken from another research the source of data must be referred.

GOOD GRAPHICAL PRACTICES

- Any graphical representation must to have:
 - Title
 - Definitions of axes
 - Units of measurements for each axe (if it is applied)
 - Legend (if it is applied)
- A good graphical representation must be as self-explanatory as possible!

GOOD GRAPHICAL PRACTICES

- The aim of a graphical representation is to transmit an information
- When drawing a graphical representation try to answer to the following question: Which is the aim of the graphical representation?
- Medical data must be represented graphically in a such a way in which to be useful for understanding the clinical phenomena
- Notice to:
 - The color composition (do not use color background)
 - The font size (it is suppose to be readable)

ONE QUALITATIVE VARIABLE: FREQUENCY TABLES

- Data are sort ascending
- The absolute frequency of each value is
- The distinct values and associated frequencies are included into a table :
 - Absolute frequency: the total amount of occurrences of one variable
 - Relative frequency = the absolute frequency divided by the total amount of occurrences

ONE QUALITATIVE VARIABLE: FREQUENCY TABLES

- Could contains the following types of frequencies:
 - ❑ Absolute frequency
 - ❑ Cumulative absolute frequency (ascending / descending)
 - ❑ Relative frequency
 - ❑ Cumulative relative frequency (ascending / descending)
- Microsoft Excel:
 - ❑ COUNTIF
 - ❑ Tabele Pivot
 - [Data - Pivot Table and Pivot Chart Report ...]

PICTORIAL SUMMARIES: ONE VARIABLE

■ Bar chart:

- The choice of interval is important (an unfortunate choice of intervals can change the apparent pattern of the distribution).
 - Enough intervals should be used so that the pattern will be minimally altering the beginning and ending positions.
- The choice of number, width, and starting points of intervals arise from the user's judgment (they should be considered carefully before forming the chart).

PICTORIAL SUMMARIES: ONE VARIABLE

- Histogram (quantitative variable):
 - Appears like the bar chart but differs in that the number of observations lying in an interval is represented by the area of a rectangular (or bar) rather than its height.
 - If all intervals are of equal width, the histogram is no different from the bar chart except cosmetically (no blank space between bars).
- Pie Chart (qualitative variable on frequency tables):
 - Represents proportions rather than amounts.
 - Its main use is to visualize the relative prevalence of the phenomena.
 - Has the advantage of avoiding the illustration of sequence that sometimes is implied by the bars charts.

PICTORIAL SUMMARIES: ONE VARIABLE

■ Line Chart:

- The main use: to convey information similar to a bar chart but for intervals that form a sequence of time or order of events from left to right.
- Relationship of a Line Chart to a Probability Distribution: as the sample size increases and the width of the intervals decreases, the line chart of a sample distribution approaches the picture of its probability distribution.

ONE QUALITATIVE VARIABLE: FREQUENCY TABLES

Absolute frequency

Relative frequency

Diagnosis	No. patients	Percent (%)
Asphyxia at birth	527	26.1
Obstetrical injuries	92	4.6
Septic status	7	0.3
Pneumonia	181	9.0
Diarrhea	8	0.4
Congenital malformations	598	29.6
Other causes	606	30.0
Total	2019	100

ONE QUALITATIVE VARIABLE: FREQUENCY

TABLES

The sum of absolute frequencies of all values in the series that are less than or equal to x/n

The sum of absolute frequencies of all values in the series that are less than or equal to x

Diagnosis	f_a	f_r	f_a cumulat \uparrow	f_r cumulat \uparrow
Asphyxia at birth	527	26.10	527	26.10
Obstetrical injuries	92	4.56	619	30.66
Septic status	7	0.35	626	31.01
Pneumonia	181	8.96	807	39.97
Diarrhea	8	0.40	815	40.37
Congenital malformations	598	29.62	1413	69.99
Other causes	606	30.01	2019	100
Total	2019	100		

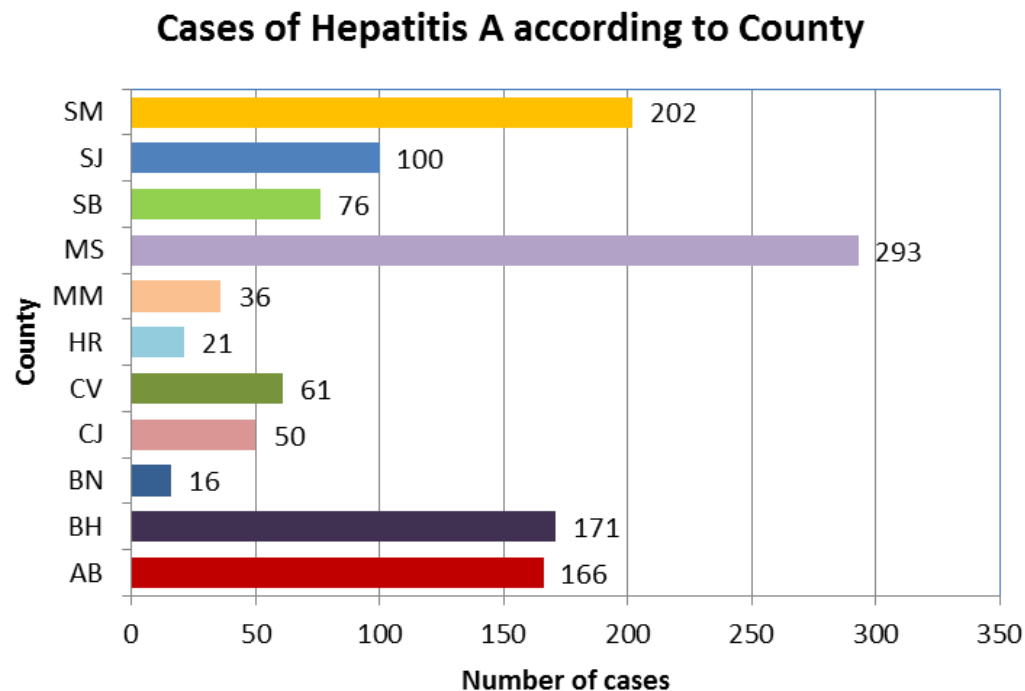
ONE QUALITATIVE VARIABLE: FREQUENCY TABLES

- Let's have the following incubation time expressed in days for a infectious diseases: 5, 6, 7, 7, 8, 8, 5, 7, 8, 7. Which of the following values correspond to the ascending cumulative relative frequency of 0.7?
 - A. 8
 - B. 6
 - C. 5
 - D. 7
 - E. No right answer

GRAPHICAL REPRESENTATIONS

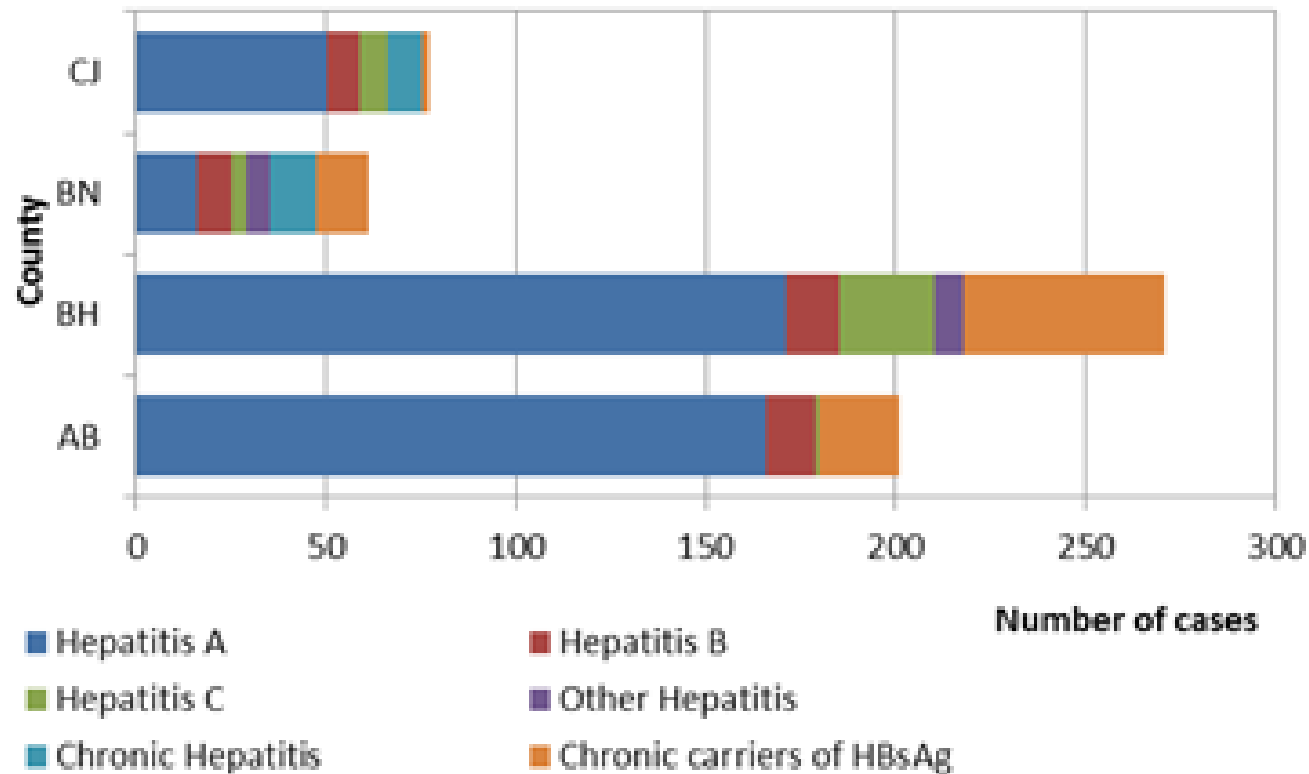
BAR CHART

- A bar graph is composed of discrete bars that represent different categories of data. The length or height of the bar is equal to the quantity within that category of data. Bar graphs are best used to compare values across categories.



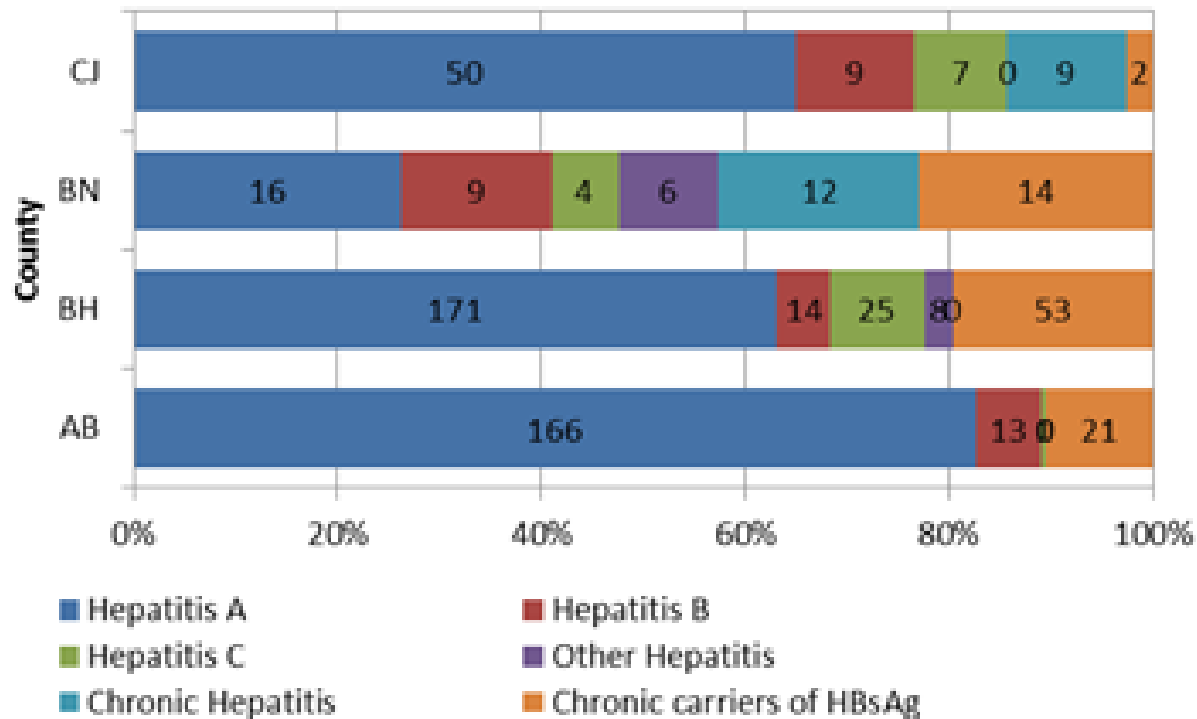
STACKED BAR

	AB	BH	BN	CJ
Hepatitis A	166	171	16	50
Hepatitis B	13	14	9	9
Hepatitis C	1	25	4	7
Other Hepatitis	0	8	6	0
Chronic Hepatitis	0	0	12	9
Chronic carriers of HBsAg	21	53	14	2



100% STACKED BAR

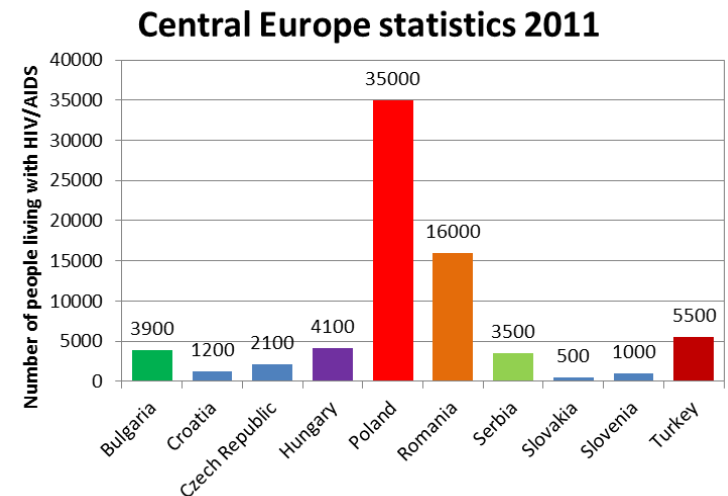
	AB	BH	BN	CJ
Hepatitis A	166	171	16	50
Hepatitis B	13	14	9	9
Hepatitis C	1	25	4	7
Other Hepatitis	0	8	6	0
Chronic Hepatitis	0	0	12	9
Chronic carriers of HBsAg	21	53	14	2



COLUMN

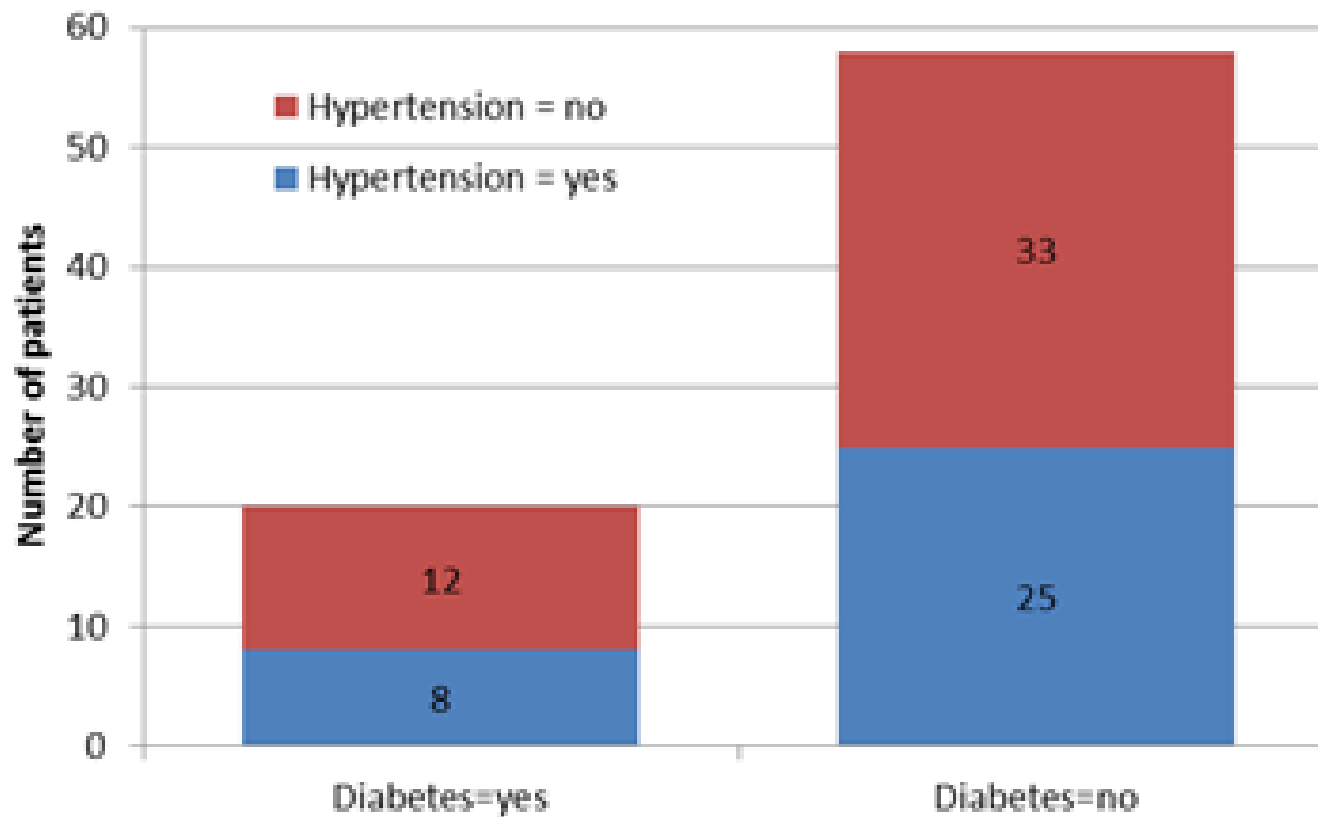
- A column graph is composed of discrete columns that represent different categories of data. The length or height of the column is equal to the quantity within that category of data. Similar with the Bar graphs, Columns graphs are used to compare values across categories.

Country	People living with HIV/AIDS 2011
Bulgaria	3900
Croatia	1200
Czech Republic	2100
Hungary	4100
Poland	35000
Romania	16000
Serbia	3500
Slovakia	500
Slovenia	1000
Turkey	5500



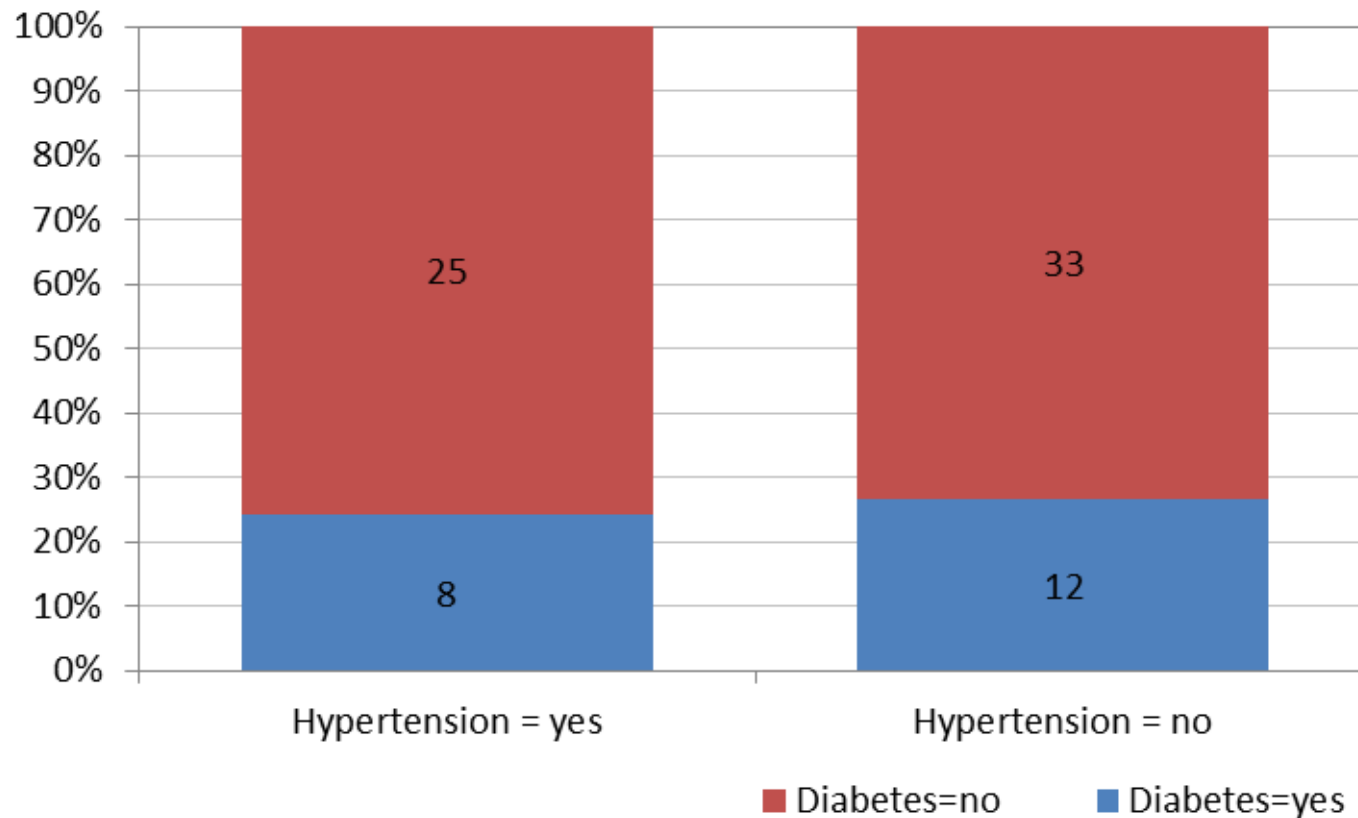
STACKED COLUMN

	Diabetes=yes	Diabetes=no
Hypertension = yes	8	25
Hypertension = no	12	33



100% STACKED COLUMN

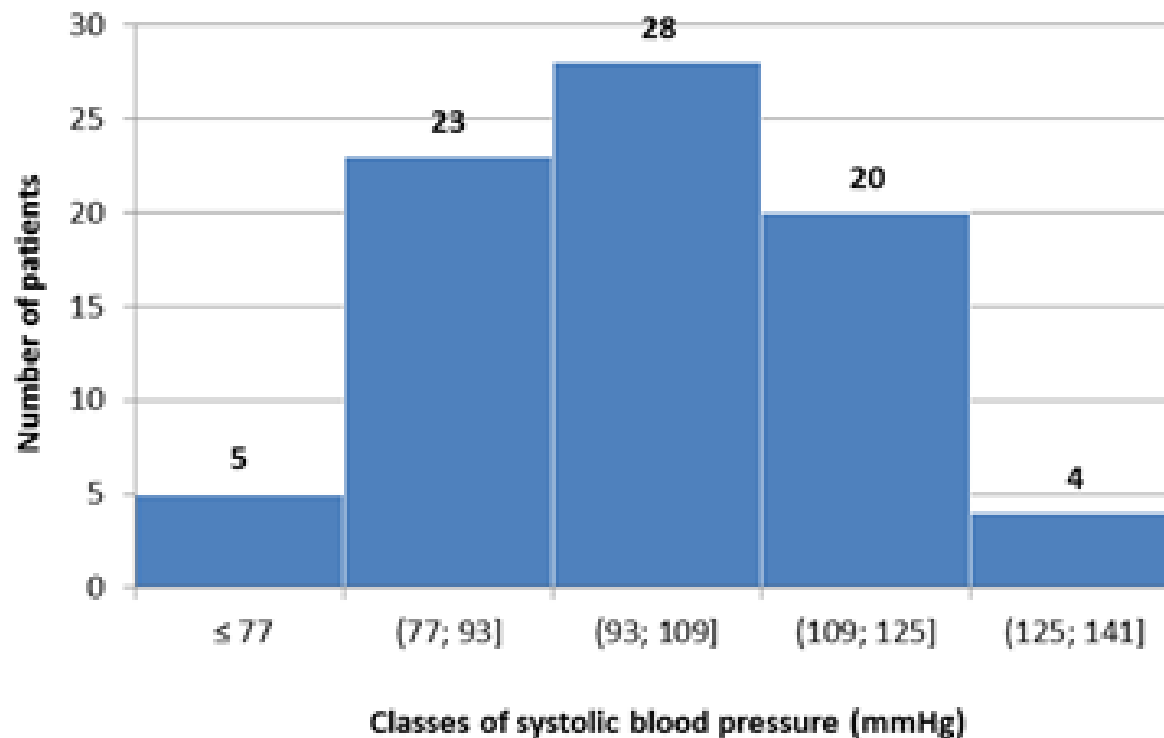
	Diabetes=yes	Diabetes=no
Hypertension = yes	8	25
Hypertension = no	12	33



HISTOGRAM

<i>Classes of frequency</i>	<i>Frequency</i>
≤ 77	5
(77; 93]	23
(93; 109]	28
(109; 125]	20
(125; 141]	4

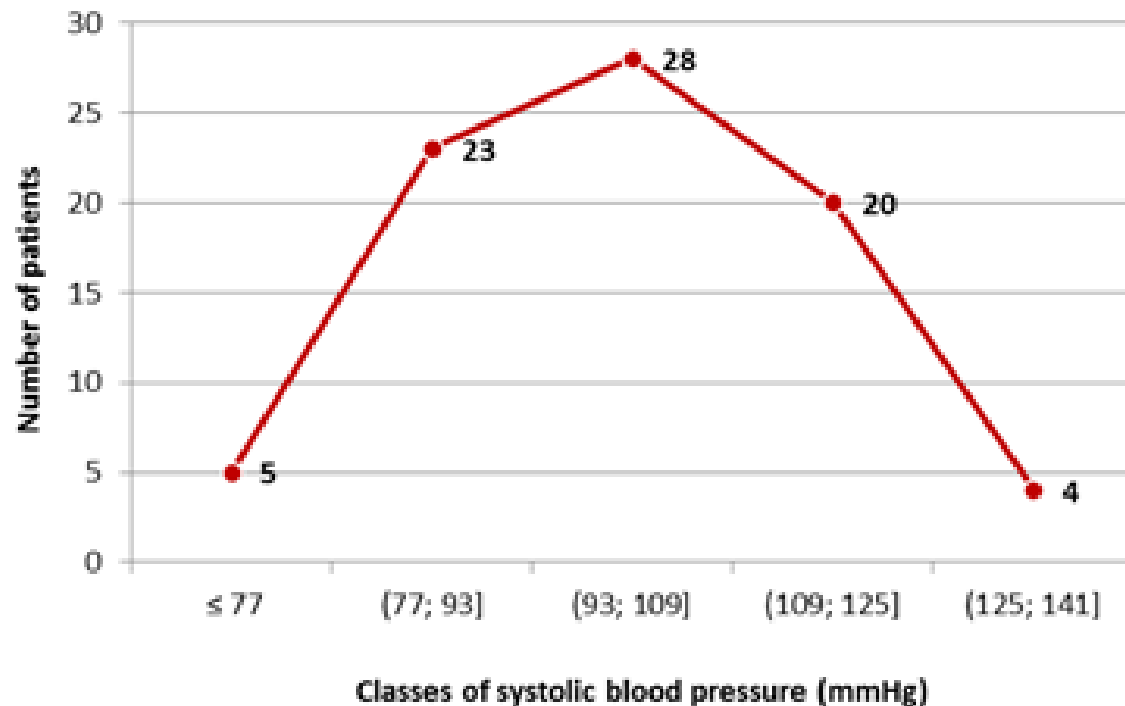
Histogram of systolic blood pressure



HISTOGRAM

<i>Classes of frequency</i>	<i>Frequency</i>
≤ 77	5
(77; 93]	23
(93; 109]	28
(109; 125]	20
(125; 141]	4

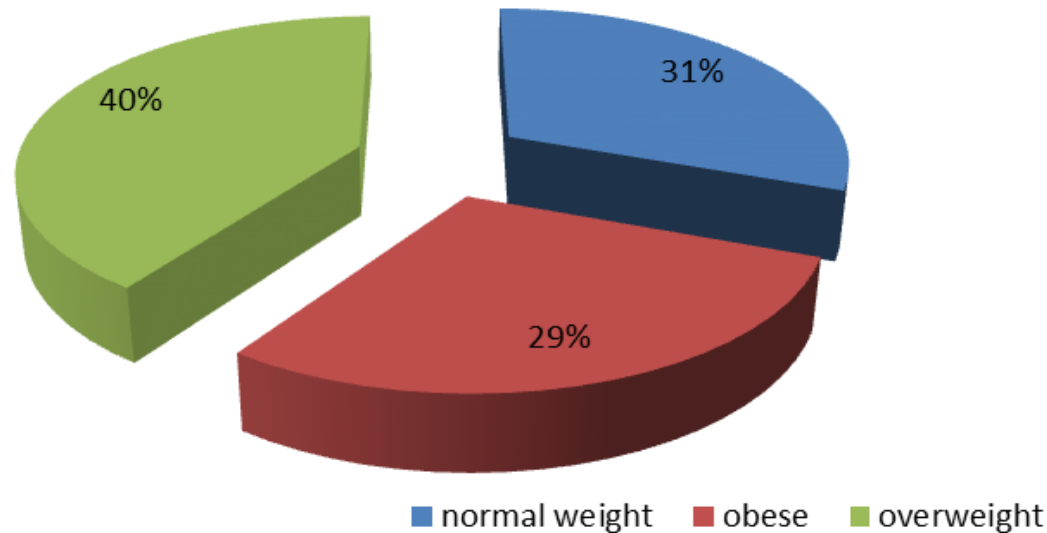
Bell shape graph of systolic blood pressure



PIE

- A Pie chart is a circular chart used to compare parts of the whole and to look to frequencies of a qualitative variable

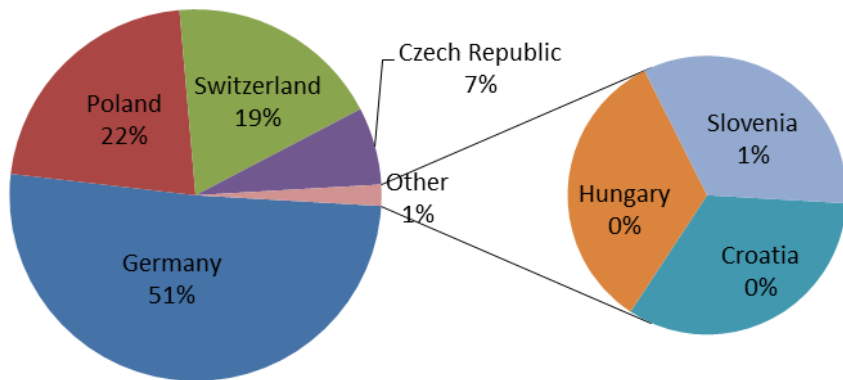
Body Mass Index classes



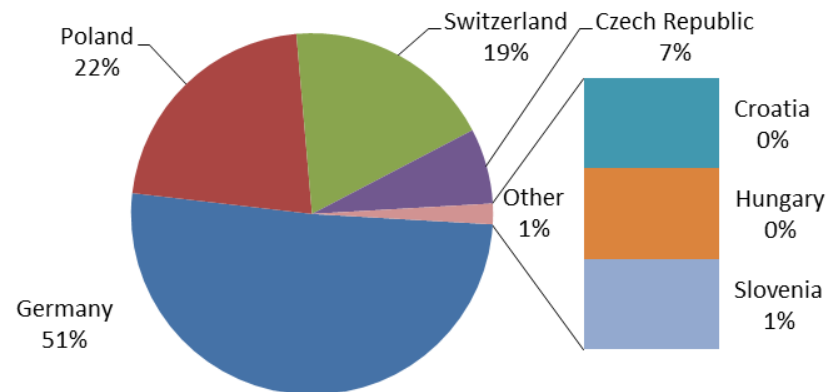
PIE OF PIE / PIE OF BAR

Country	No of cases of measles: 2012
Germany	166
Poland	71
Switzerland	61
Czech Republic	22
Croatia	2
Hungary	2
Slovenia	2

**Distribution of Measles in Central Europe:
2012**



**Distribution of Measles in Central Europe:
2012**

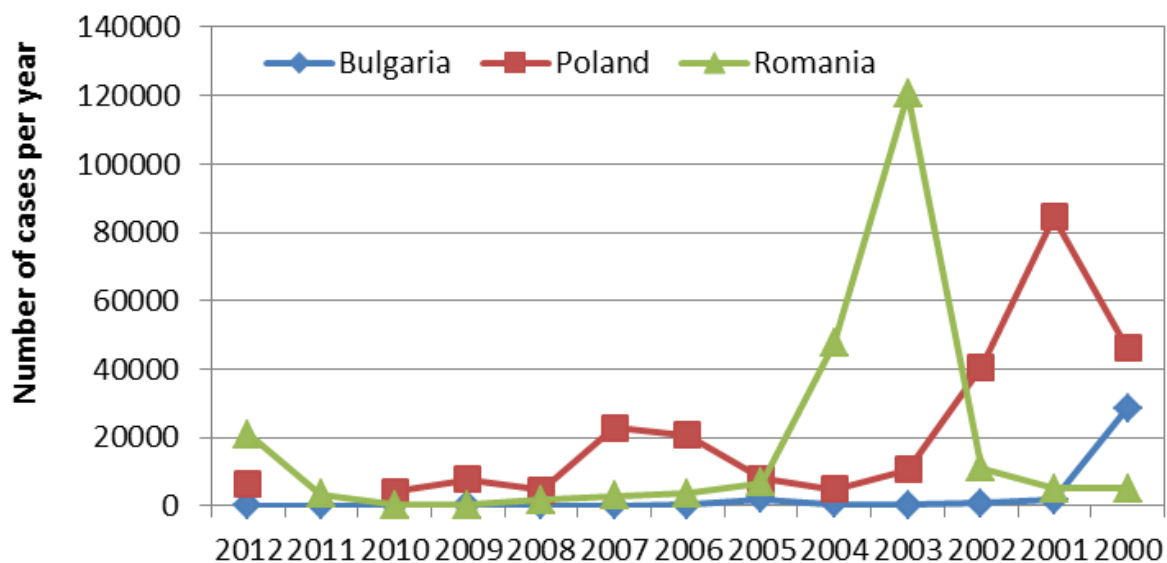


LINE

- Line graphs provide information on the relation between two variables and are used to illustrate trends over the time.

Countryname	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000
Bulgaria	18	41	39	44	58	88	247	1968	474	252	805	1655	28449
Poland	6263	0	4196	7586	4598	22890	20668	7946	4857	10588	40518	84419	46181
Romania	20812	3494	350	343	1746	2958	3563	6801	47444	120377	11079	5076	5125

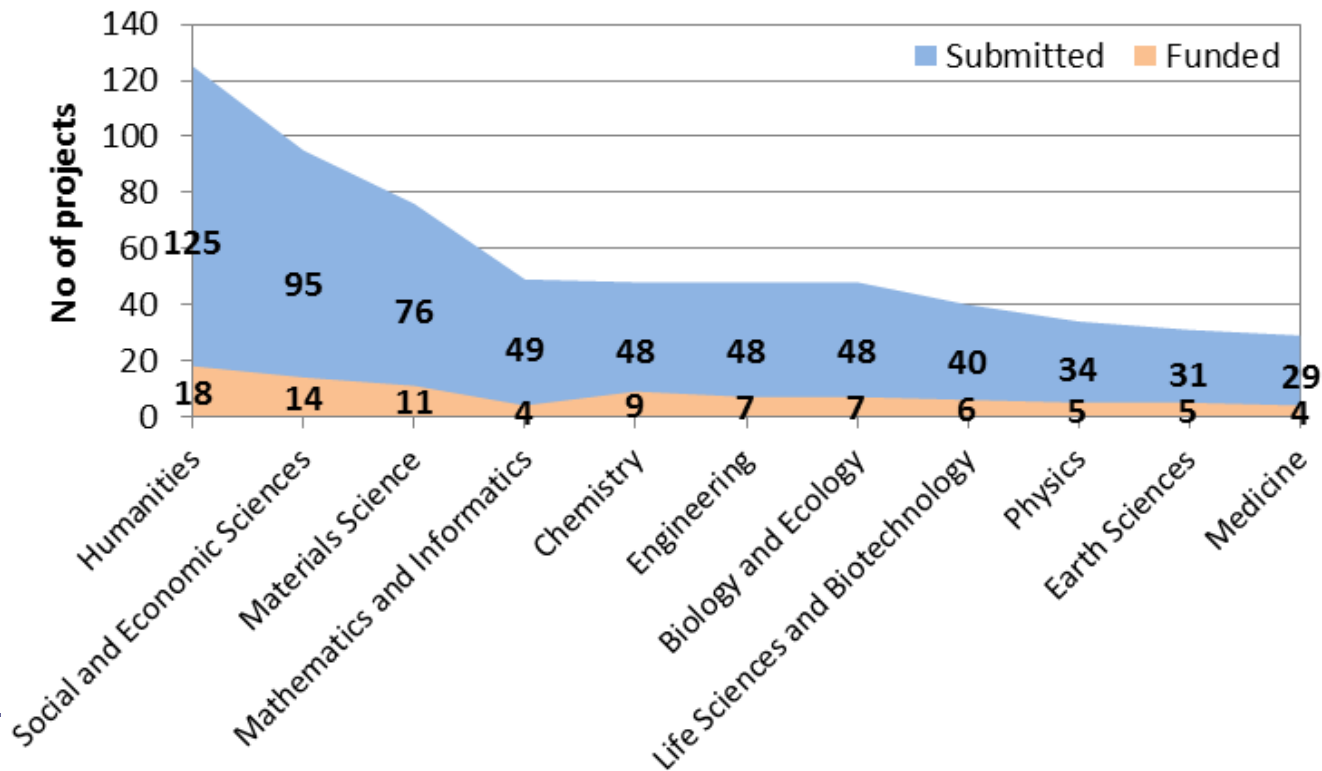
Trends of Rubella



AREA

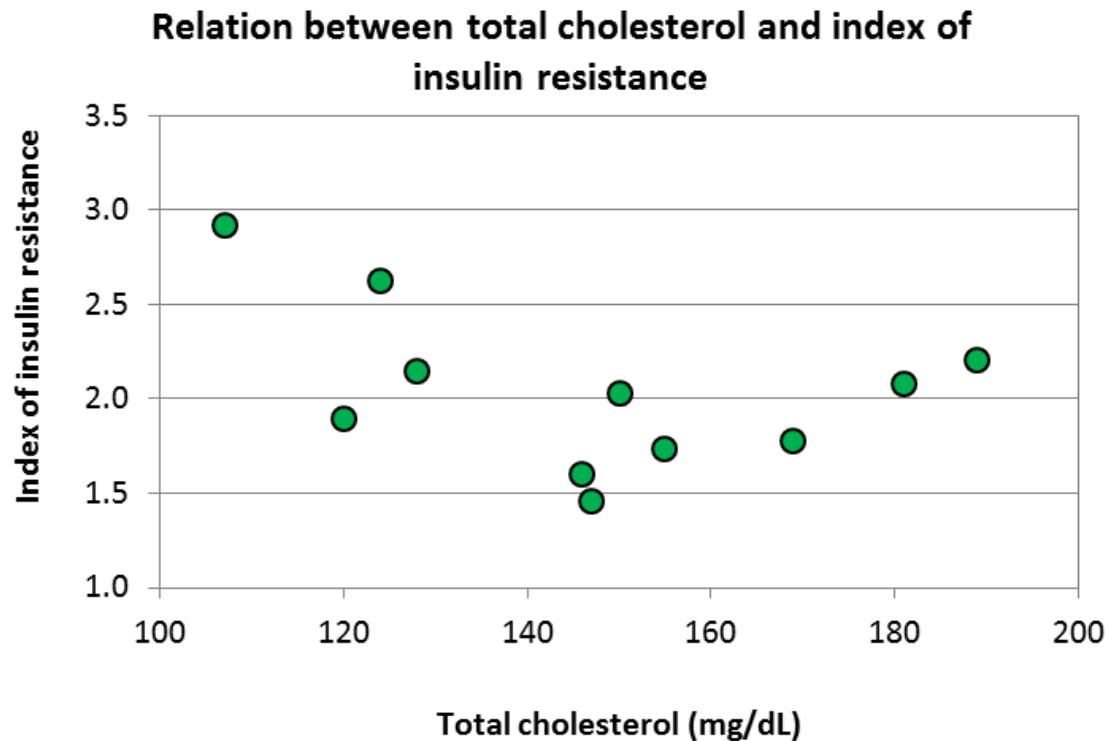
Research Domain	Submitted	Funded
Humanities	125	18
Social and Economic Sciences	95	14
Materials Science	76	11
Mathematics and Informatics	49	4
Chemistry	48	9
Engineering	48	7
Biology and Ecology	48	7
Life Sciences and Biotechnology	40	6
Physics	34	5
Earth Sciences	31	5
Medicine	29	4

PostDoctoral Research Grants Competition 2012



SCATTER

Total Cholesterol (mg/dL)	Index of insulin resistance
181	2.08
146	1.60
155	1.73
107	2.92
128	2.14
120	1.90
150	2.03
169	1.77
147	1.46
189	2.21
124	2.62



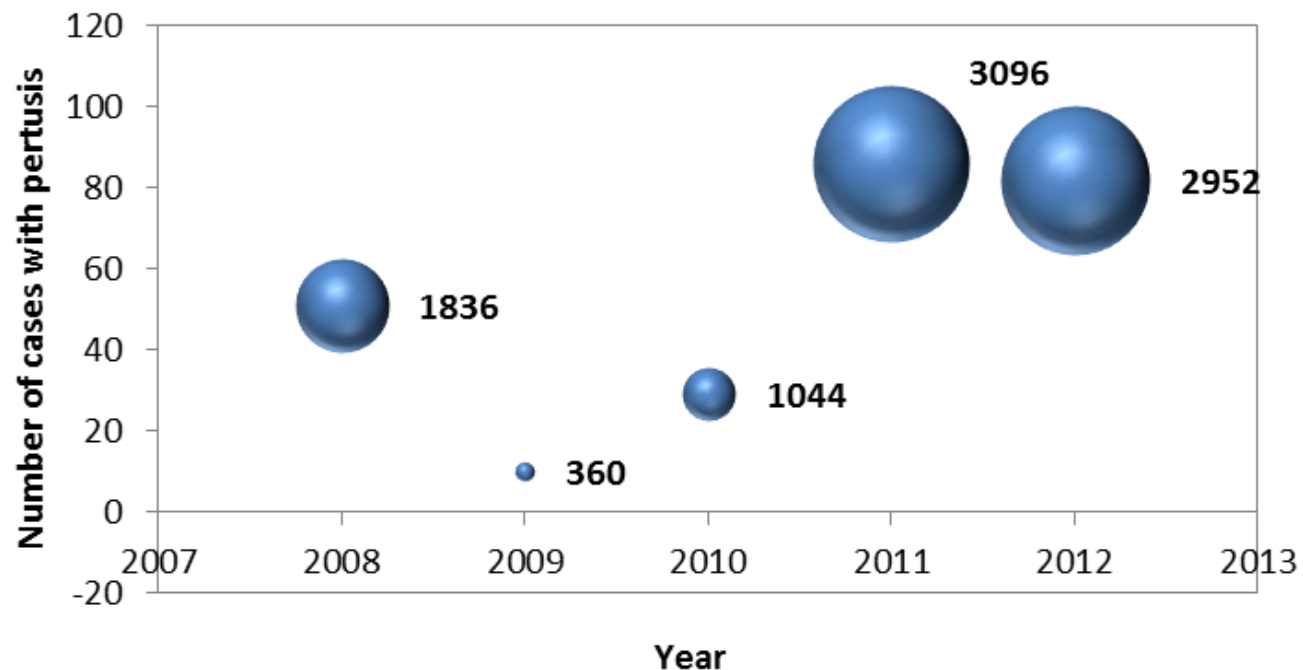
BUBBLE

- A bubble chart is a type of chart that displays three dimensions of data. Each entity with its triplet (v_1, v_2, v_3) of associated data is plotted as a disk that expresses two of the v_i values through the disk's xy location and the third through its size. The size of the bubble (data marker) indicates the value of the third selected variable.

BUBBLE

Disease	Pertussis	Average antibiotics costs
2012	82	2952
2011	86	3096
2010	29	1044
2009	10	360
2008	51	1836

Average antibiotics costs



ONE QUALITATIVE VARIABLE: FREQUENCY TABLES

- Let have the following incubation time expressed in days for a infectious diseases: 5, 6, 7, 7, 8, 8, 5, 7, 8, 7. Which of the following values correspond to the ascending cumulative relative frequency of 0.7?

Value	f_a	f_r	f_a cc	f_r cc
5	2	0.20	2	0.20
6	1	0.10	3	0.30
7	4	0.40	7	0.70
8	3	0.30	10	1
Total	10	1		

TWO QUALITATIVE VARIABLES: CONTINGENCY TABLE

	TBC+	TBC-	Total
Gender=F	2	10	12
Gender=M	24	54	78
Total	26	64	90

n QUALITATIVE VARIABLES: FREQUENCY TABLES

Table 1. Distribution of pulmonary pathologies associated with silicosis

Grade of silicosis	BrC	BPOC	Emphysema	CPC	TBC	Total
I	12	20	0	0	14	46
I/II	1	5	1	1	1	9
II	3	7	1	1	7	19
II/III	0	1	0	0	0	1
III	0	3	0	0	4	7
Total	16	36	2	2	26	82
BrC = chronic bronchitis; BPOC = broncho-pneumonitis chronic obstructive; CPC = chronic pulmonary heart; TBC = pulmonary tuberculosis						

STATISTICAL SUMMARY BY EXAMPLE

- Gökhan Açıköz, Murat İnanç Cengiz, İlker Keskiner, Şereften Açıköz, Murat Can, and Aydan Açıköz. Correlation of Hepatitis C Antibody Levels in Gingival Crevicular Fluid and Saliva of Hepatitis C Seropositive Hemodialysis Patients. International Journal of Dentistry 2009; Article ID 247121.

Table 1: Crosstabulation of HCV antibodies Immunoreactivity in Gingival Crevicular fluid and Saliva, Kappa = 0.426; $p < .001$.

		Gingival Crevicular fluid							
		Positive		Gray Zone		Negative		Total	
		n	%	n	%	n	%	n	%
Saliva	Positive	2	5.1			3	7.7	5	12.8
	Gray Zone			3	7.7			3	7.7
	Negative	4	10.3	1	2.6	26	66.7	31	79.5
	Total	6	15.4	4	10.3	29	74.4	39	100

GOOD TABLES PRACTICES: SUMMARY!

- Tables:
 - Capture: information concisely and display it efficiently
 - Provide information at any desired level of detail and precision
 - Number tables consecutively in the order of their first citation in the text and supply a brief title for each
 - Give each column a short or an abbreviated heading. Authors should place explanatory matter in footnotes, not in the heading
 - Explain all nonstandard abbreviations in footnotes
 - Identify statistical measures of variations
 - If you use data from another published or unpublished source, obtain permission and acknowledge that source fully

GOOD GRAPHIC PRACTICES: SUMMARY!

- Figures should be made as self-explanatory as possible.
- Titles and detailed explanations belong in the legends-not on the illustrations themselves.
- Figures should be numbered consecutively according to the order in which they have been cited in the text.
- If a figure has been published previously, acknowledge the original source and obtain written permission from the copyright holder to reproduce the figure.
- Explain clearly in the legend each symbols, arrows, numbers, or letters used in a figure.
- Avoid 3D graphical representations!

SUMMARIZING DATA - GRAPHS

- **SCATTER PLOT:**
 - two continuous numerical values
- **BAR GRAPH:**
 - qualitative variables
- **LINE GRAPH:** one quantitative variable
- **HISTOGRAM:** one continuous variable
- **PIE CHART:** one/two qualitative variables