WEEK 2

SUMMARIZING DATA: TABLES AND DIAGRAMS

SUMMARIZING DATA

- As the quantity of information grows, obtaining an overall “picture” of what is happening becomes increasingly difficult…

A) TABLES

- A table is an orderly arrangement, usually of numbers or words in rows and columns, which exhibits a set of facts in a distinct and comprehensive way.

A table should;

Include a concise, informative and unambiguously defined title.

Give a brief heading for each row and column.

Include the units of measurement.

Give the number of items on which any summary measure (e.g. a percentage) is based.

Provide a summary statistic (e.g. the mean) a measure of dispersion (e.g. a std. deviation)

Only give necessary information

Table 2. Estimated mean career duration of Arabian horses that joined the competitions between 2006 and 2012 in Turkey. Results were obtained from Kaplan Meier analysis for each categories of factors investigated.
Tablo 2. Türkiye'de 2006-2012 yılları arasında yarışlara katılan Arap atları için hesaplanan ortalama kariyer süreleri. Sonuçlar her bir değişken için Kaplan Meier analizi yardımıyla hesaplanmıştır.

Variable	Category	N	%	Estimated (t)	Std. Error	95% CI		p
						L.B.	U.B	
Age group	<=3.5 years	2954	88.1%	23.22	0.39	22.46	23.99	<0.001
	>3.5 years	400	11.9%	19.38	0.89	17.64	21.12	
Ownership status	Sole-ownership	3234	96.4%	22.8	0.36	22.08	23.51	0.143
	Joint-ownership	120	3.6%	21.02	1.92	17.27	24.77	
Sex	Male	1784	53.2%	26.99	0.56	25.9	28.08	<0.001
	Female	1570	46.8%	18.03	0.4	17.25	18.81	
Type of stud farm	State	1152	34.3%	27.15	0.67	25.84	28.46	<0.001
	Private	2202	65.7%	20.32	0.4	19.53	21.11	
Racing on a single type of track surface	Yes	264	7.9%	14.83	0.98	12.9	16.76	<0.001
	No	3090	92.1%	23.34	0.38	22.6	24.07	
Earnings	<11000 $*	1125	33.5%	10.63	0.18	10.28	10.97	<0.001
	11001-35700 $	1130	33.7%	16.52	0.26	16.01	17.04	
	>35700 $	1099	32.8%	39.71	0.69	38.36	41.05	

*(t)=month., * US$.*

SOURCE: Ozen & Gurcan (2016). Determination of factors affecting the length of racing career of Arabian horses in Turkey. Ankara Univ Vet Fak Der 63, 303-309

B) DIAGRAMS

- A diagram is a graphic representation of data and may take several forms. It is often easier to discern important *patterns* from a diagram rather than a table, even though the latter may give more precise numerical information.

- Diagrams are most useful when we want to convey information quickly, and they should serve as an adjunct to more formal statistical analysis.

Statistical diagrams serve two main purposes:

(1) the presentation of statistical information in articles and other reports, when it may be felt that the reader will appreciate a simple, evocative display.

(2) as a private aid to statistical analysis. The statistician will often have recourse to diagrams to gain insight into the structure of the data and to check assumptions which might be made in an analysis.

Ref: Aviva Petrie & Paul Watson (2013). Statistics for Veterinary and Animal Science, Third Edition, Wiley-Blackwell.

NOTES ON DIAGRAMS !

Keep it simple, avoid unnecessary ‘vanity’ (make a simple pie chart, histogram or bar chart).

Include a concise + informative defined title.

Label all axes, segments and bars, if necessary using a legend or key showing the meaning of the different symbols used.

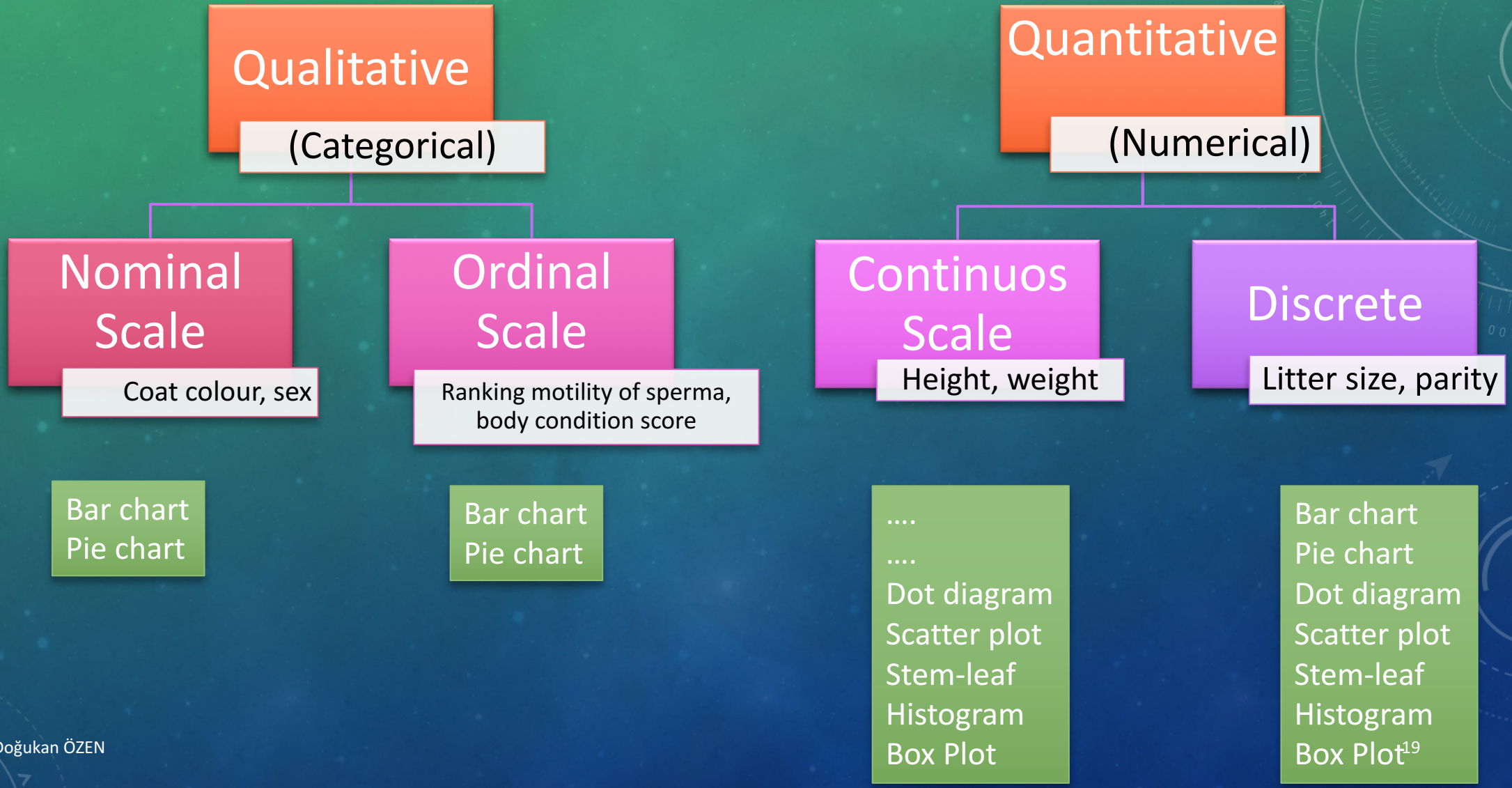
Present the units, the numbers on which summary measures are based, and measures of variability where appropriate.

Avoid exaggerating the scale on an axis

Ensure that the method of display conveys all the relevant information (e.g. pairing).

B) DIAGRAMS

SELECTING CORRECT TYPE OF DIAGRAM



BAR CHART

-Displays a vertical bar for each category

-The length of each bar, which should be of constant width, depicts the number or percentage of individuals belonging to that category.

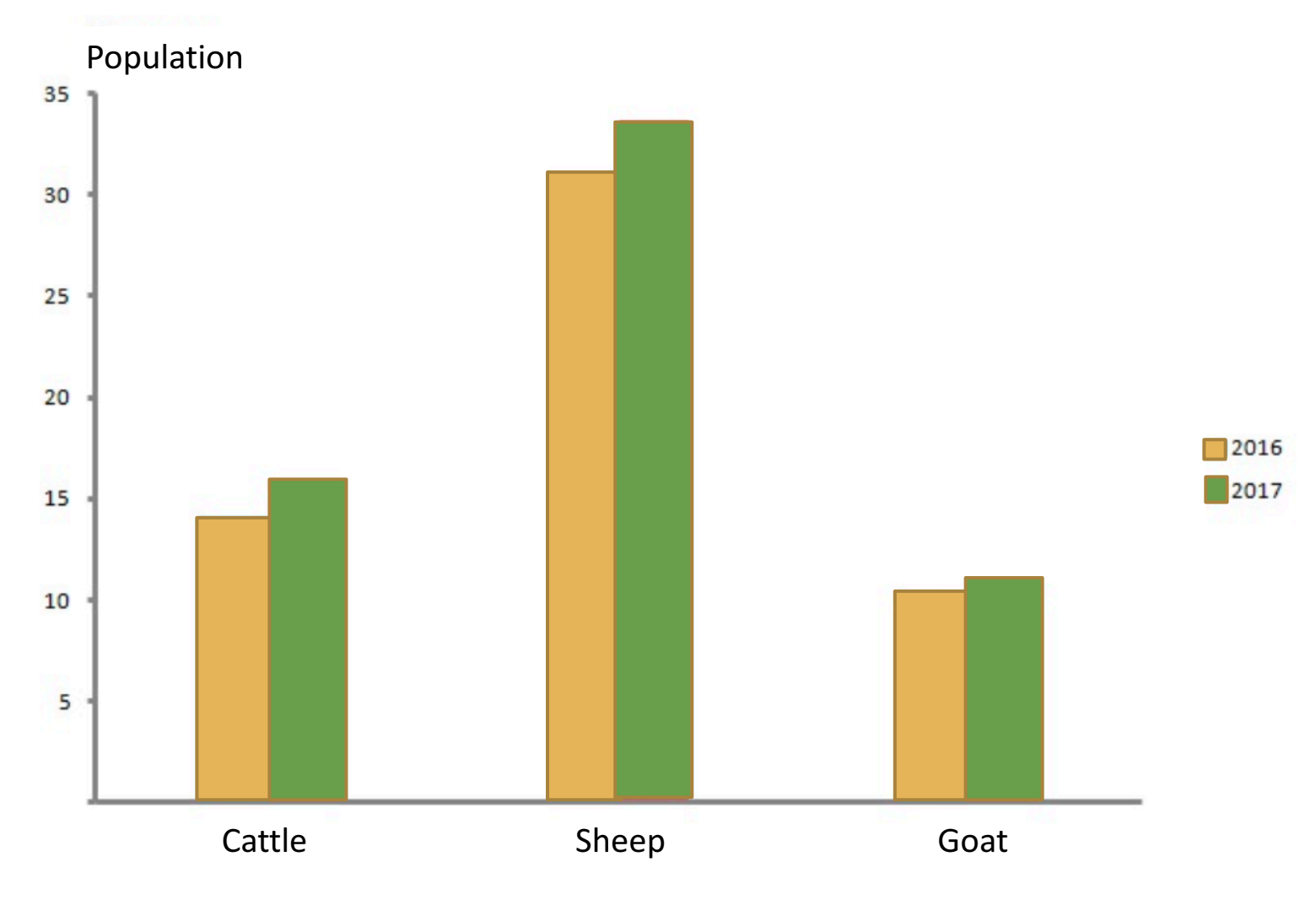
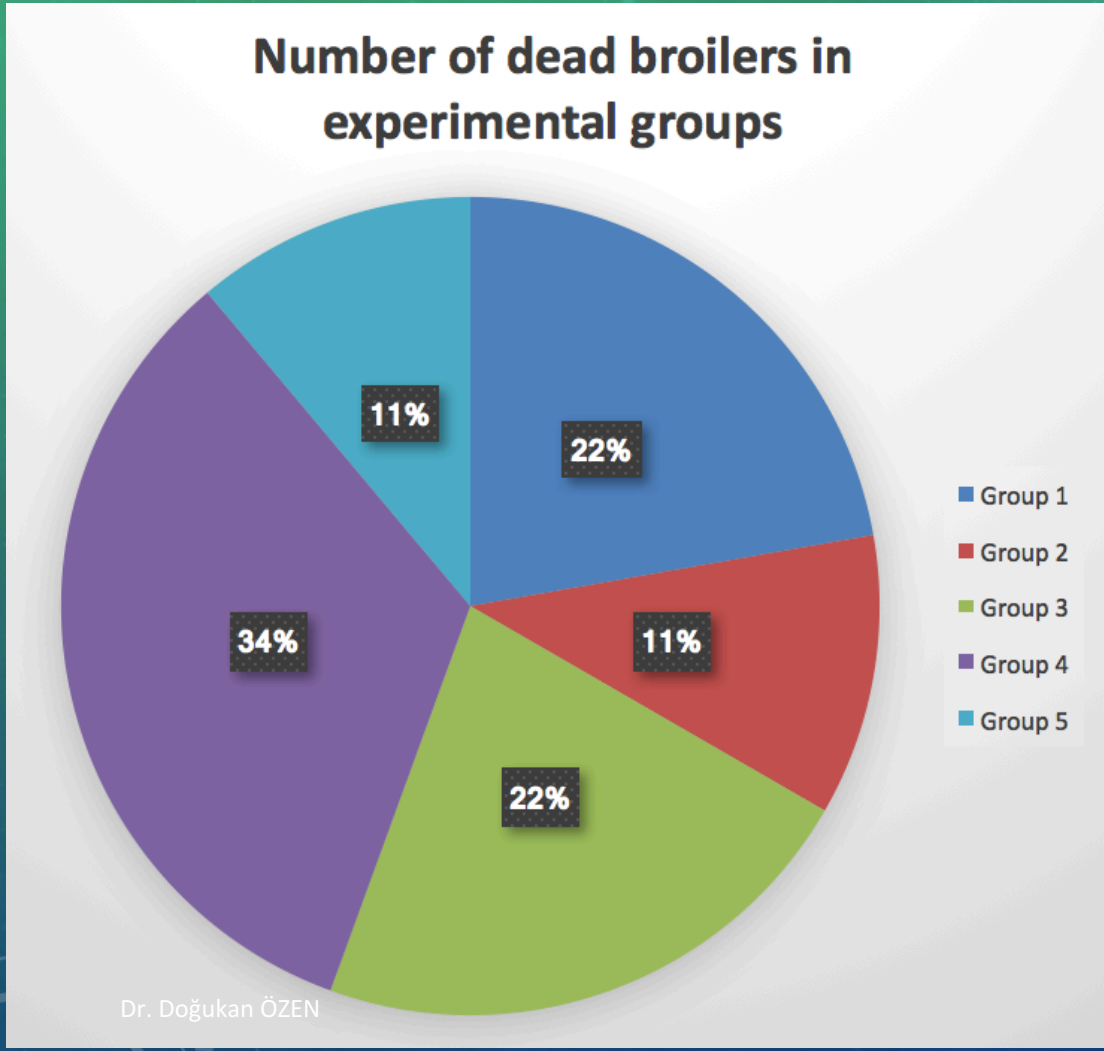


Figure 1. …….. Appropriate title…

PIE CHART



A circle divided into segments with each segment portraying a different category

LINE DIAGRAM

To express the change in some quantity over a period of time

The natural method here is a graph in which points, representing the values of the quantity at successive times, are joined by a series of straight-line segments

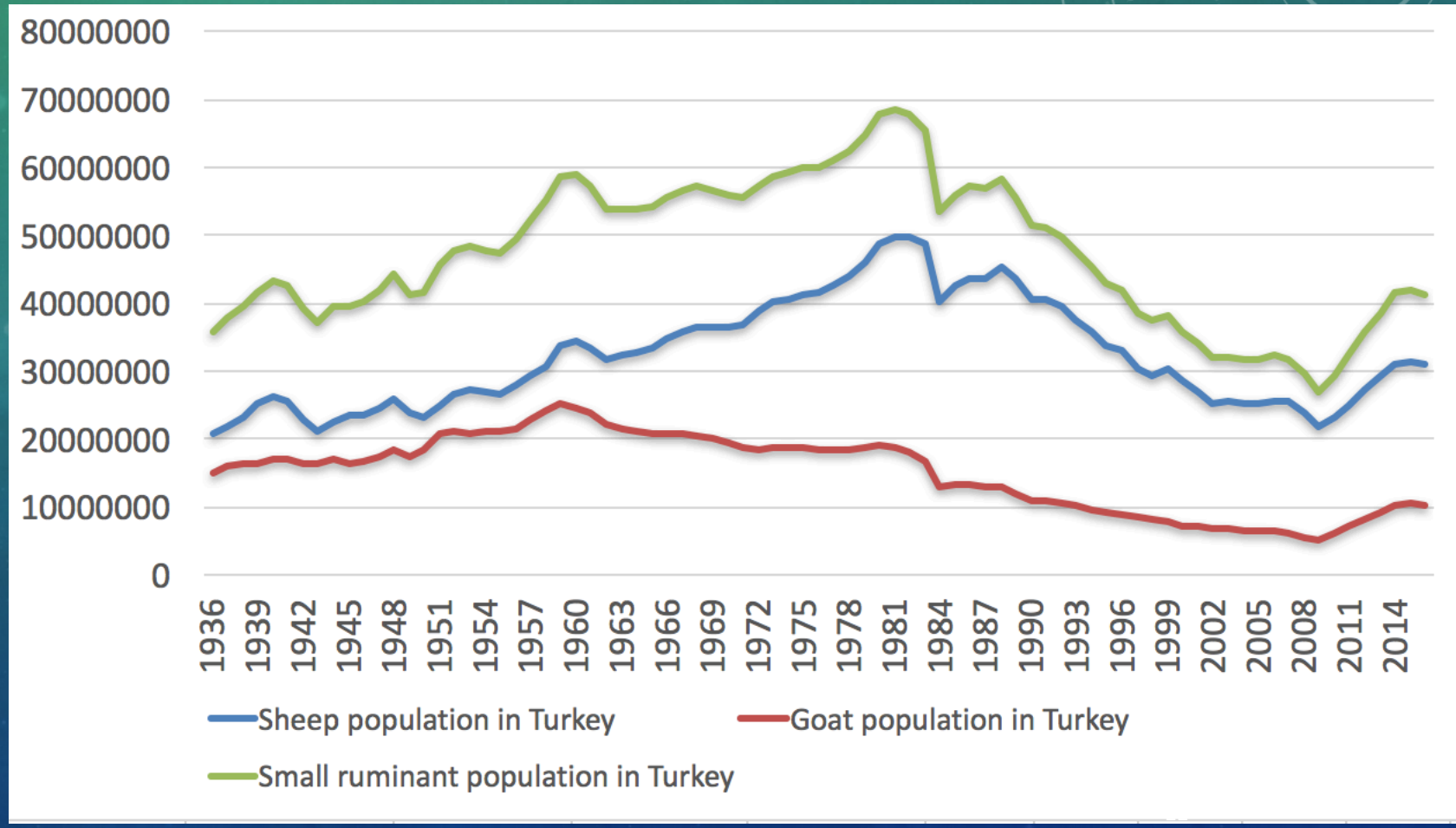


Figure 1. Small ruminant population in Turkey between 1936 and 2014

STEM AND LEAF DIAGRAM

```
Bodyweight-  Stem-and-Leaf Plot
 Frequency     Stem &  Leaf
     1.00        3  .   8
     3.00        4  .   001
     2.00        4  .   23
     5.00        4  .   44455
     3.00        4  .   667
     8.00        4  .   88888889
     8.00        5  .   00000111
     5.00        5  .   23333
     6.00        5  .   444445
     5.00        5  .   66777
     2.00        5  .   89
     3.00        6  .   001
     2.00        6  .   22

 Stem width:      10.00
 Each leaf:        1 case(s)
```

Stems	Leaves
7	69
8	00125699
9	12446

Observation = 92, in a sample of 15 test scores

- It resembles a histogram that has been turned on its side. Each vertical rectangle of the histogram is replaced by a row of numbers that represent the relevant observations.
- The stem is the core value of the observation (e.g. the unit value before the decimal place) and the leaves are represented by a sequence of ordered single digits, one for each observation, that follow the core value (e.g. the first decimal place)

HISTOGRAM

A two-dimensional diagram used to express the distribution of individual objects or measurements into different categories in which usually the horizontal axis represents the unit of measurement of the variable of interest, with each class interval being clearly delineated.

It is appropriate form of diagram when the number of observations is large the original data may be grouped into a frequency distribution table

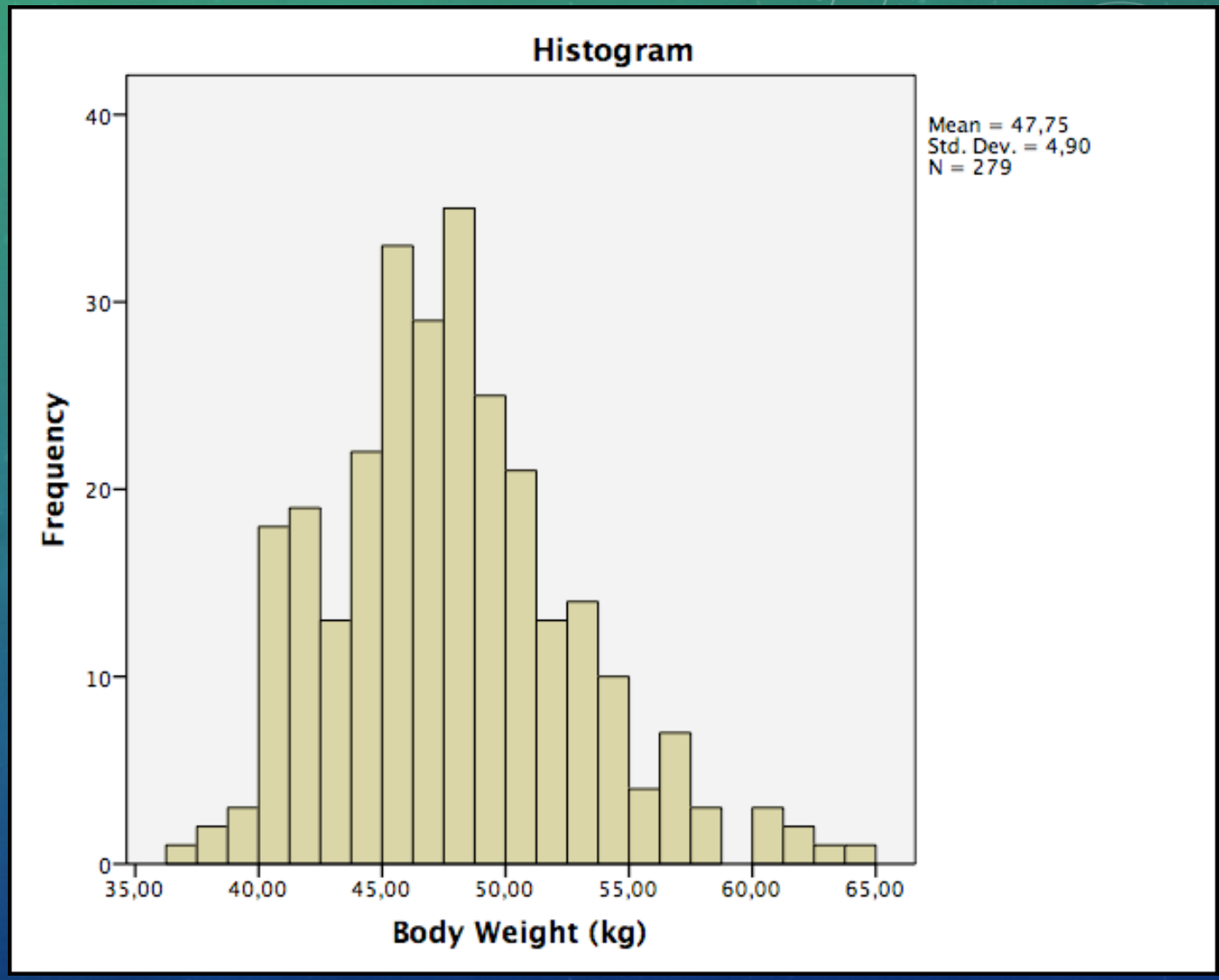
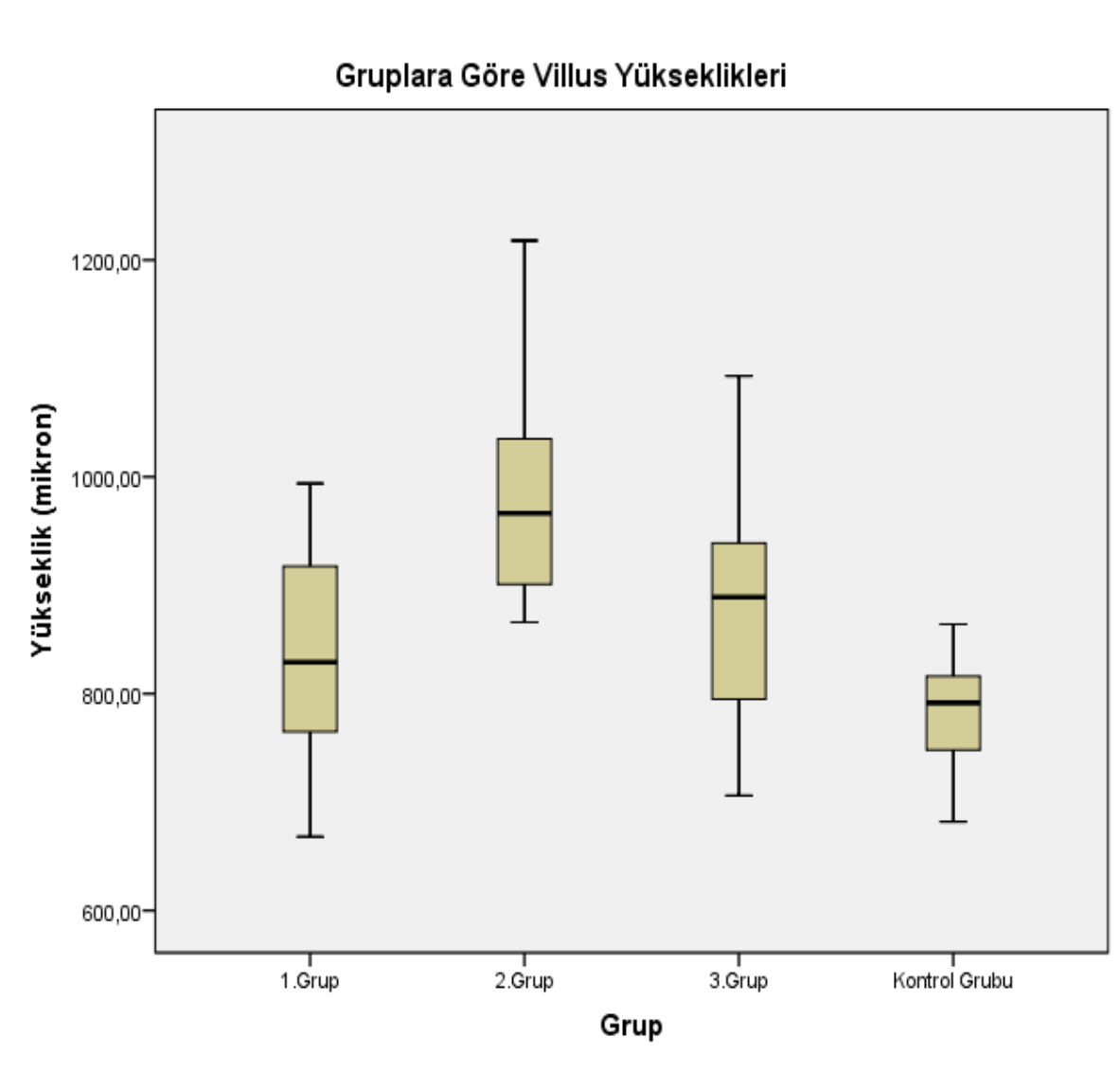
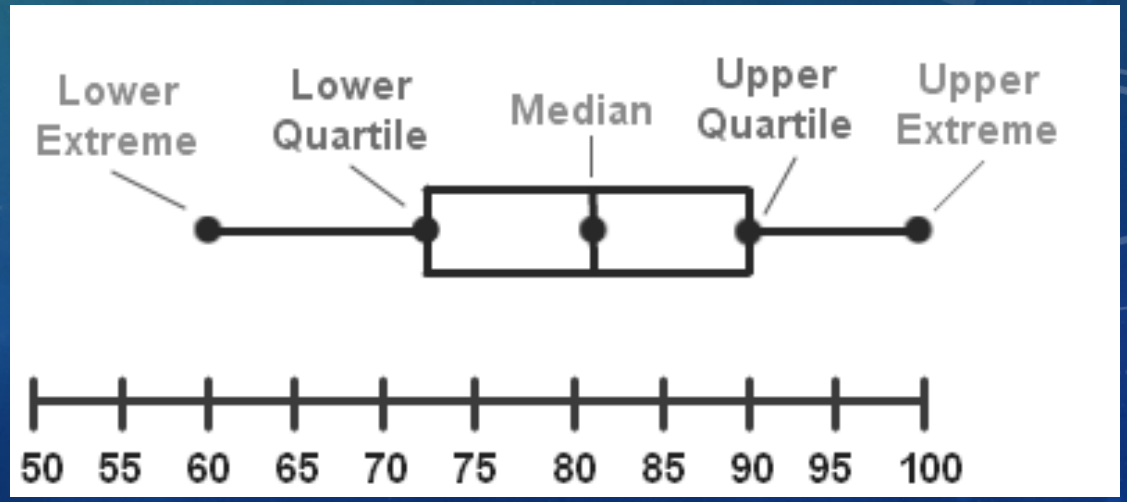


Figure 1. Body weight of 279 Awasi sheep.

BOX AND WHISKER PLOT



- The diagram comprises a box with horizontal limits defining the upper and lower quartiles and representing the interquartile range, enclosing the central 50% of the observations, with the median marked by a horizontal line within the box.
- The box-and-whisker plot is particularly useful when a number of data sets are to be compared in a single diagram



SCATTER DIAGRAM

- To express the relationship between two measurements, in a situation where they occur in pairs.

- An effective way of presenting data when we are interested in examining the relationship between two variables which may be numerical or ordinal.

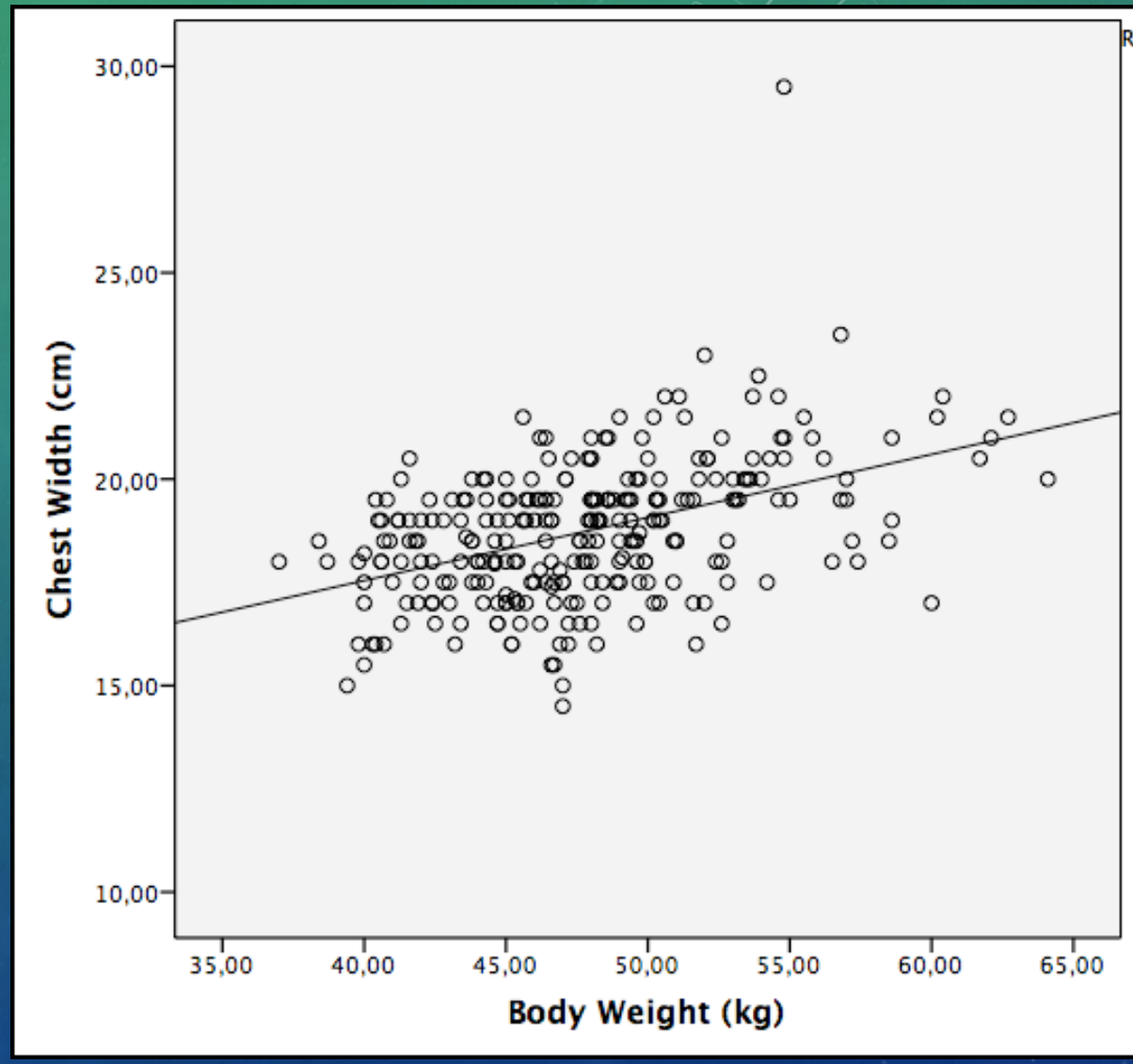
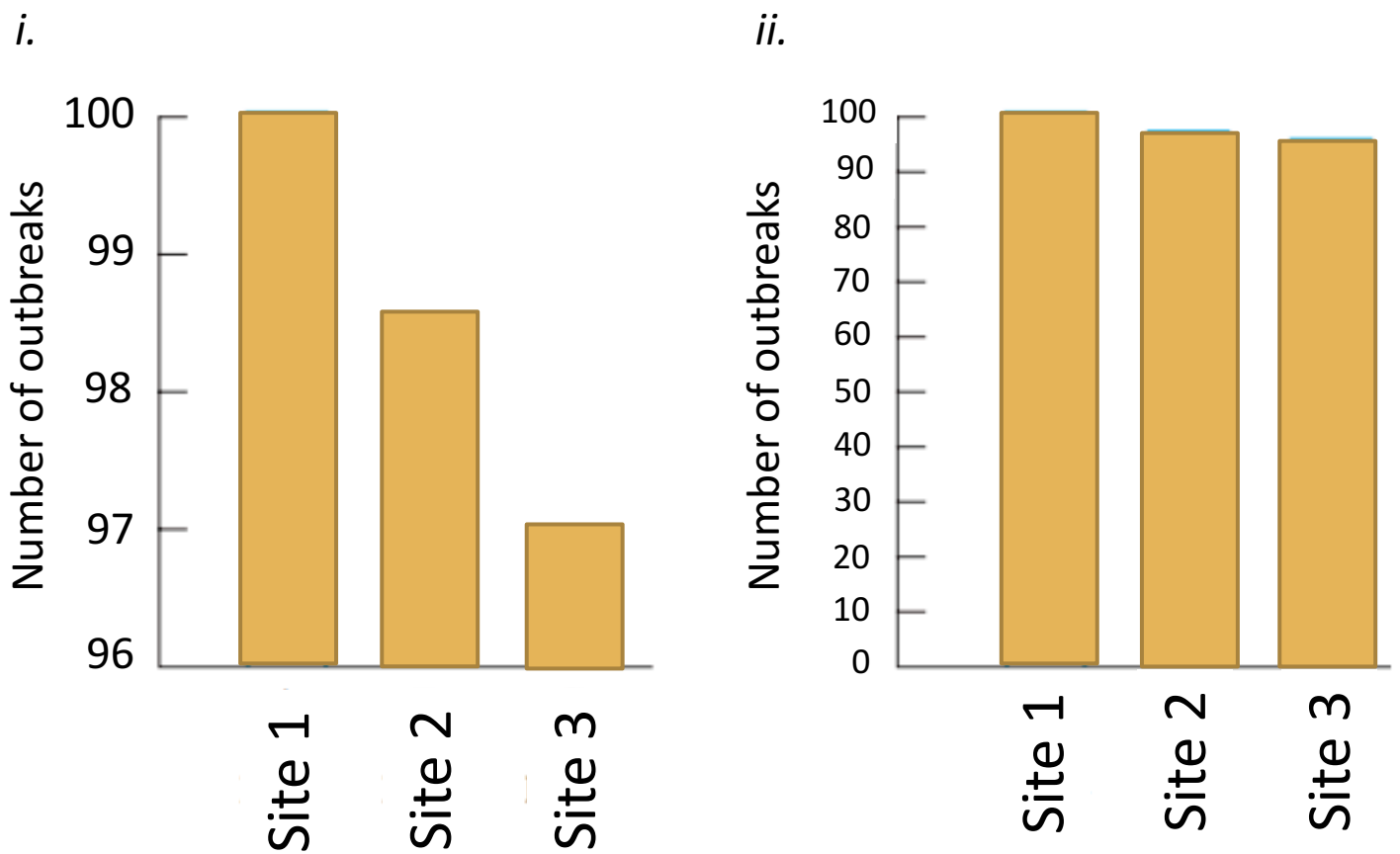


Figure 1. Relationship of chest width and body weight measures in Awasi sheeps

MISLEADING GRAPHS !!

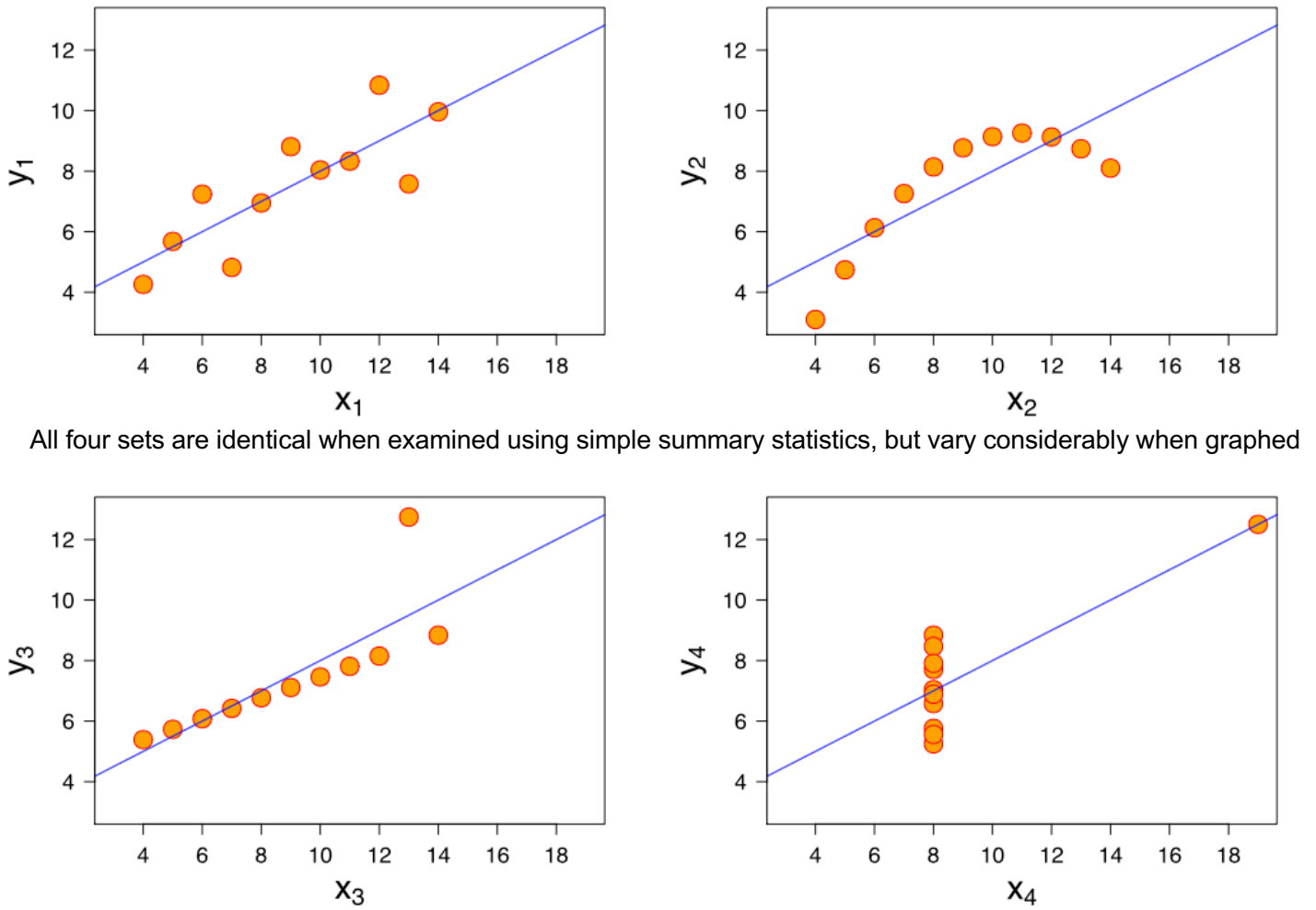
Graphs should be constructed so that they accurately portray the essential characteristics of data



REMEMBER ! IT IS ALSO IMPORTANT TO CHECK YOUR DATA GRAPHICALLY BEFORE ANALYZING IT

ANSCOMBE'S QUARTET

Anscombe's quartet comprises four datasets that have nearly identical simple descriptive statistics, yet appear very different when graphed. Each dataset consists of eleven (x, y) points. They were constructed in 1973 by the statistician Francis Anscombe to demonstrate both the importance of graphing data before analyzing it and the effect of outliers on statistical properties. He described the article as being intended to counter the impression among statisticians that "numerical calculations are exact, but graphs are rough."



All four sets are identical when examined using simple summary statistics, but vary considerably when graphed

Ref: Anscombe, F. J. (1973). "Graphs in Statistical Analysis". *American Statistician*. 27 (1): 17–21. and the related page in Wikipedia.