Data Description – P&P Questions

1. Data types

a. A set of 10 hypothetical patient records from a large database is presented in Table 1. Patients with a diabetes value of 1 have type-II diabetes and patients with a diabetes value of 0 do not have type-II diabetes. For each of the following variables, assign them to one of the following scales: nominal, ordinal, interval, or ratio:
(a) Name
(b) Age
(c) Gender
(d) Blood group
(e) Weight (kg)
(f) Height (m)
(g) Systolic blood pressure (mmHg)
(h) Diastolic blood pressure (mmHg)
(i) Diabetes

TABLE 1 Table of Patient Records

Name	Age	Gender	Blood Group	Weight (kg)	Height (m)	Systolic Blood Pressure (mmHg)	Diastolic Blood Pressure (mmHg)	Diabetes
P. Lee	35	Female	A Rh+	50	1.52	68	112	0
R. Jones	52	Male	O Rh−	115	1.77	110	154	1
J. Smith	45	Male	O Rh+	96	1.83	88	136	0
A. Patel	70	Female	O Rh−	41	1.55	76	125	0
M. Owen	24	Male	A Rh−	79	1.82	65	105	0
S. Green	43	Male	O Rh−	109	1.89	114	159	1
N. Cook	68	Male	A Rh+	73	1.76	108	136	0
W. Hands	77	Female	O Rh−	104	1.71	107	145	1
P. Rice	45	Female	O Rh+	64	1.74	101	132	0
F. Marsh	28	Male	O Rh+	136	1.78	121	165	1

Source: [MSD]

2. Distribution visualisation

a. The number of absentees in a class over a period of 24 days were:
0, 3, 1, 2, 1, 0, 4, 0, 1, 1, 2, 3, 1, 0, 0, 2, 4, 6, 4, 2, 1, 0, 1, 1
By first drawing up a tally chart obtain a frequency table and draw a bar chart.
Source: [US]

b. The total scores in a series of basketball matches were:
215, 224, 182, 200, 229, 219,
209, 217, 195, 162, 210, 213,
204, 208, 197, 192, 187, 213
Construct a stem-and leaf diagram to represent the above data.
Source: [US]

c. The shoe sizes of the members of a football team are:
10, 10, 8, 11, 10, 9, 9, 10, 11, 9, 10
Represent the data on a bar chart.
Source: [US]

d. Using the data in Table 2, create a histogram of Sale Price ($) using the following intervals: 0 to less than 250, 250 to less than 500, 500 to less than 750, and 750 to less than 1000.

TABLE 2 Retail Transaction Data Set

Customer	Store	Product Category	Product Description	Sale Price ($)	Profit ($)
B. March	New York, NY	Laptop	DR2984	950	190
B. March	New York, NY	Printer	FW288	350	105
B. March	New York, NY	Scanner	BW9338	400	100
J. Bain	New York, NY	Scanner	BW9443	500	125
T. Goss	Washington, DC	Printer	FW199	200	60
T. Goss	Washington, DC	Scanner	BW39339	550	140
L. Nye	New York, NY	Desktop	LR21	600	60
L. Nye	New York, NY	Printer	FW299	300	90
S. Cann	Washington, DC	Desktop	LR21	600	60
E. Sims	Washington, DC	Laptop	DR2983	700	140
P. Judd	New York, NY	Desktop	LR22	700	70
P. Judd	New York, NY	Scanner	FJ3999	200	50
G. Hinton	Washington, DC	Laptop	DR2983	700	140
G. Hinton	Washington, DC	Desktop	LR21	600	60
G. Hinton	Washington, DC	Printer	FW288	350	105
G. Hinton	Washington, DC	Scanner	BW9443	500	125
H. Fu	New York, NY	Desktop	ZX88	450	45
H. Taylor	New York, NY	Scanner	BW9338	400	100

Source: [MSD]

e. The masses (in g to the nearest g) of a random collection of offcuts taken from the floor of a carpenter's shop are summarised below:

0-19	20-39	40-59	60-99
4	17	12	6

Display the data using a histogram.
Source: [US]

f. The marks gained in an examination are summarised below.

0-29	30-49	50-69	70-99
4	12	37	14

Represent the data using a histogram.
Source: [US]

g. The following data are the numbers of deaths of army officers caused by horse kicks, for the Prussian Army during the period 1875 to 1894. In order of size the numbers are:

3, 4, 5, 5, 6, 6, 7, 8, 9, 9, 10,
11, 11, 11, 12, 14, 15, 15, 17, 18

Find the range and interquartile range.
Illustrate the data using a box–whisker diagram.

Source: [US]

h. One year the numbers of academic staff (including part-time staff) in the various departments of the University of Essex (a small, friendly university) were as follows:

19.0, 15.7, 25.3, 28.0, 15.0, 10.0,
12.0, 10.3, 22.0, 24.8, 13.8, 25.9,
23.0, 21.3, 12.0, 11.0, 23.0

Find the range and interquartile range.
Illustrate the data using a box–whisker diagram.

Source: [US]

3. Distribution parameters

a. A random sample of seven runner beans have lengths (in cm, to the nearest cm) given as 28, 31, 24, 33, 28, 32, 30. Find the value of S (the standard deviation).
Source: [US]

b. In an experiment, a cupful of cold water is poured into a kettle and the time take for the water to boil is noted. The experiment was conducted six times giving the following results (in seconds): 125, 134, 118, 143, 128, 131. Find the value of S^2 (the variance).
Source: [US]

c. A random sample of 15 observations has sample mean 11.2 and sample variance 13.4. One observation of 21.2 is judged to be unreliable. Find the sample mean of the remaining 14 observations.
Source: [US]

d. Calculate the following statistics for the variable Age (from Table 3):
(a) Mode
(b) Median
(c) Mean
(d) Range
(e) Variance
(f) Standard deviation
Source: [MSD]

(g) Calculate any additional required values and plot a box-and-whisker diagram, indicating the 5 important points.

TABLE 3 Table with Variables *Name* and *Age*

Name	Age
P. Lee	35
R. Jones	52
J. Smith	45
A. Patel	70
M. Owen	24
S. Green	43
N. Cook	68
W. Hands	77
P. Rice	45
F. Marsh	28