

Assessment Brief Proforma

1. Module number	<i>SET09120</i>
2. Module title	Data Analytics
3. Module leader	<i>Thomas Methven</i>
4. Tutor with responsibility for this Assessment Student's first point of contact	<i>Thomas Methven</i>
5. Assessment	<i>Coursework</i>
6. Weighting	<i>35% of module assessment.</i>
7. Size and/or time limits for assessment	<i>The page limit is 8 pages, more details in following specification</i>
8. Deadline of submission	<i>Friday 23rd October 2020, 15:00 UK Time</i>
9. Arrangements for submission	Your Coursework must be submitted to the coursework submission link provided on Moodle.
10. Assessment Regulations	All assessments are subject to the University Regulations
11. The requirements for the assessment	<i>See handout</i>
12. Special instructions	
13. Return of work and feedback	<i>Feedback will be via rubric and/or free text on Moodle where appropriate</i>
14. Assessment criteria	<i>See Handout</i>

Data Visualisation Coursework (SET09120)

Thomas Methven

School of Computing, Edinburgh Napier University, Edinburgh
t.methven@napier.ac.uk

Abstract. In this coursework, you will have to explore a dataset using visualisation, find relevant patterns, creating visualisations showing these patterns, and then explain what these patterns actually mean. You will also be required to comment on the good and bad aspects of a visualisation someone else has created, and sketch a better alternative way of showing that data.

Finally, you will need to submit a report that contains your findings and visualisations. This report will be in the LNCS Style, and will be no longer than eight (8) pages, with only four (4) pages of text, and four (4) of images. The template for this is available on your Moodle page.

More information about this coursework is presented in this document, which you should make sure you read in detail before beginning. If you fail to follow the required style, you might miss out on marks!

Keywords: You won't need these!

1 Overview

Participants from four differently skilled groups have been asked to take part in some sort of throwing sport. What sport it is exactly has not been explained to you, but you know several measurements have been taken for each participant's throw. Each participant only throws once. Their gender and age has been recorded, alongside the wind direction at the time of their throw, as well as the results of the throw – max height, distance, throw angle, and offset from a straight throw. Finally, two unknown and unrelated aspects of their throw have been scored, and recorded. Bear in mind that this information has been anonymised so you, for example, do not know how the scores are calculated, or how the groups were selected.

You have been provided a dataset which contains 20,000 of these measurements, and have been asked to identify any trends, relationships, or other interesting information or patterns in it. You are also expected to understand what this data means, so you can state these patterns in a sensible way. In other words, rather than saying 'P throws further than N' you will be expected to say 'Professionals, in general, throw a further distance than Novices'. This example is not one of the expected patterns.

In order to help with this, the following metadata information has been provided:

Table 1. A table showing the metadata for the dataset you have been provided.

Column Name	Data Description
Group	N, E, V, P (Novice, Experienced, Very Exp, Professional)
Gender	M or F
Age	Numeric Value
Wind.Direction	N, E, S, W
Height	Numeric Value (in metres)
Distance	Numeric Value (in metres)
Angle	Numeric Value (in degrees)
Offset	Numeric Value (in centimetres)
Score.1	Numeric Value (out of 100)
Score.2	Numeric Value (out of 100)

There are also two classes of outliers in the data that you are expected to find, highlight, and report on. Remember: each class of outlier will have more than one row of data associated with it. In addition, if these outliers are making your other visualisations more difficult to understand, you are expected to remove said data before creating the visualisations. More information on how to report these is provided below.

Finally, you have been given a visualisation someone else has created. This is provided in **Fig. 1** of this document, and shows the average female height of different countries. With reference to the figure, you should critically evaluate the way in which the data has been visually encoded, highlighting any problems. You will also sketch and describe a more appropriate way to visualise this data, giving justification for your design choices based on the visualisation theory provided during the lectures.

2 Report

Your report **must** use the Springer's Lecture Notes in Computer Science (LNCS) template. Templates for Latex and Word are provided on the Moodle page for this module. The page limit is **eight (8) pages**, of which the first four contain only text, and the final four will contain your visualisations with appropriate captions. Deviation from either the template or the page limits will affect your mark. This document is in LNCS format. Your report should include the following sections:

1. *Abstract* – This section, if read alone, should give the reader a clear, albeit brief, picture of what you have done and found.
2. *Description of the outliers found* – Through exploration of the data, you should find outliers in the data. In this section you should discuss why you believe these data are outliers, and provide suitable visualisations as evidence. These should be included in your four page image allowance, and you

should use appropriate internal referencing to make them easy to find. You should also state if you removed said outliers, and if so why.

Note: There are two classes of outlier in the data.

3. *Description of the relationships found* – Through your exploration of the data, you should find some interesting relationships between some of the attributes. In this section you should say what you found and how you found them. Include visualisations as part of your four page image allowance at the end of the document, and use appropriate internal referencing to make them easy to find. **Note: There are five relationships in the data**
4. *Visualisation Evaluation* – You have been provided a visualisation someone else has created (**Fig. 1**). You are to discuss the good and bad aspects of this visualisation. In addition, you should *sketch on paper* a better way of presenting this data, and justify why it is better. Both of these parts should be supported by the visualisation theory you have learnt in the lectures. The sketch should be included as part of your four page image allowance.

The report must be separated so that all of the written sections come first, followed by all the images. This separation is so the marker can verify you have adhered to the page limits. All included figures must have a caption, and must be referred to at some point in the text, ‘e.g. See **Fig. 1**’.

If you submit a visualisation as evidence which cannot be read easily due to resolution issues, bad formatting, or any other problem, that visualisation WILL NOT be marked.

3 Submission

The only thing that should be submitted is the PDF version of your report. Do not submit Word or TEX files. The name of the PDF should be your student number. The submission point will be made available on Moodle.

Please note: The relationships have been designed to be different difficulties to find. In other words, some will be more difficult to find than others! Your work should, of course, be your own and we will check for plagiarism.

4 Marking Scheme

The coursework will be marked out of 35, and will be worth 35% of your overall grade for the SET09120 Data Analytics module. The breakdown of the marks will be as follows:

1. *Outliers Found (5 marks)* – Have you found the two classes of outliers that were in the dataset? Did you show this with the correct type of visualisation? Did you remove the outliers if appropriate?
2. *Relationships found (15 marks)* – Have you found all five relationships in the dataset? Do your included visualisations provide evidence that you found the relationship? Did you pick the most sensible visualisation as evidence?

3. *Visualisation Evaluation (10 marks)* – Have you identified issues with the visualisation? Have you justified them with the theory discussed in the lectures? Does your sketch improve on this? Is the justification well-reasoned?
4. *Quality of report (5 marks)* – Adherence to the template, quality of written English, quality of images etc.

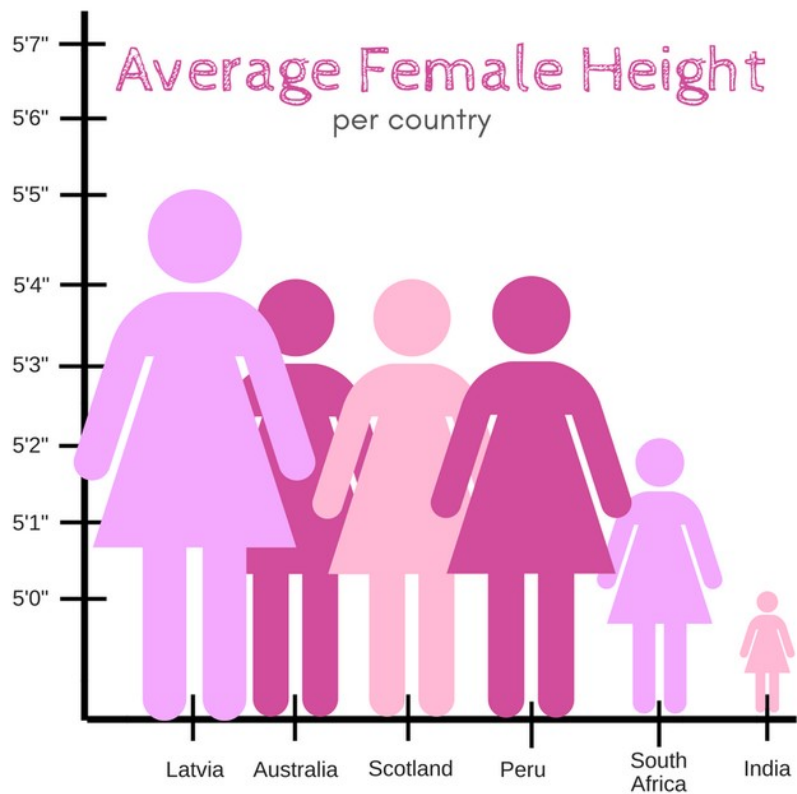


Fig. 1. This is the example visualisation you are expected to comment on. Remember, any visualisations you include in your report should be on your final four pages only, should have a caption like this, and should be referred correctly from within the text.

References (Only include this if used)

1. Author, F.: Article title. Journal 2(5), 99–110 (2016).
2. Author, F., Author, S.: Title of a proceedings paper. In: Editor, F., Editor, S. (eds.) CONFERENCE 2016, LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2016).
3. Author, F., Author, S., Author, T.: Book title. 2nd edn. Publisher, Location (1999).
4. Author, F.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1–2. Publisher, Location (2010).
5. LNCS Homepage, <http://www.springer.com/lncs>, last accessed 2016/11/21.