Data Visualization and Interpretation, DAT7002BNM, January-February 2021

Rough solution ideas for the assignment:

Important: Please note that the following are suggestions only, and that you will tell your data story as you consider appropriate.

Likewise, the lessons mentioned below are indicative, as relevant material can be found throughout the module.

Thanking you.

Before you begin - aspects to consider when working on your assignment, please:

- Arden would like you to use academic references in the assignment, that is, published books, articles, but also working papers; annual reports, company website information and internet sources like blogs can also be used

- precise referencing is crucial; all material taken from sources must be visible as such in the text, and at the end of the assignment, collected in an alphabetically sorted reference list

- some online databases for academic literature search:
* EBSCO (built into iLearn under "My Resources")
* sciencedirect.com
* ideas.repec.org (this is especially good for working papers)
* Google Scholar

- when researching a topic, it is suggested to start with a specific, detailed search term, and if nothing is found, broaden the search term

- Internet sources, such as Wikipedia, can also be used, but must be referenced, of course ...

- precise and close referencing is crucial; all material taken from sources must be visible as such in the text, with author names and year e.g. "Norten and Thannes (2017) have pointed out that ..." and at the end of the assignment, all sources are to be collected in an alphabetically sorted reference list and given as fully as possible e.g. "Norten, C. and Thannes, P. (2017): Marketing as a science, Journal of Marketing Investigations, November Issue, pages 326-345."

- please, we also urge you to abstain from what Arden UK call "collusion"; collusion happens when several students use identical/similar text in their assignments; when you write your text yourself, this will not happen :)

Format of the work required for part 1 of the assignment:

- The assignment requires you to work on a given dataset, so please follow the usual structure of empirical papers: Introduction/motivation, data description, analysis (with visuals), results, conclusions, references

- A poster simply collapses all this into one page, in most cases in landscape format (example ...); you are free how you generate this technically, so it is not compulsory to use PowerPoint

- Given that the dataset is provided, the literature research on the topic involved (please see below) will not be the focus, but some research is required for the Introduction/motivation part

- This research you will conduct as outlined in the Research Methods module, e.g. using EBSCO, and look for terms such as "email use in higher education (HE)" etc. (please see above)

- The intro also states the research question and can contain so-called anecdotal evidence on a topic; here, this looks as follows ...

- *Starting point of your analysis: Please recall the CRISP-DM framework, we need a business-oriented question, objective (decision making perspective)*

- *Assignment paper: A decision is to be made whether to provide training to employees in regard to email use; this objective will guide us as to what we look for in the data*

- The data is about email use in the UK HE sector, so we can definitely extract some information from the data to support our management team

Looking at the data:

- We need to understand the topical background of the data (email use in higher education), and the basic nature/structure of the data ...

- The data has been collected through a survey, and we have, for the most part, responses coded on a numerical scale; hence the importance of the "codebook" to understand the meaning of the values

- The codebook is the "Logfile-sav-into-gretl.txt" file, we can look at this right now, first searching for the "§" character ...

- Cross section, 1010 respondents (=observational units are the people who gave answers to the survey)

- When looking at the data (always keeping in mind our research question/objective, namely email use in HE and the need for training), we can *identify 3 groups of variables*:
* Variable group 1 ("layer 1"): Information on key issues with email usage
* Variable group 2 ("layer 2"): Information on training and its potential impact on email usage
* Variable group 3 ("layer 3"): Personal characteristics of respondents (e.g. age) i.e. "additional information"

Developing the basic storyline:

- *Please recall the idea/approach of data storytelling: Connecting parts of the data to answer a given question*

- There is seldom only one possible data story; how can we get started here?

- *Your creativity, and the research question guides you through the data: Which variables are relevant to the question, and which combinations of variables are relevant?*

- Just have a quick look at the variables "q_17_score" and "q22_score": A considerable numbe of respondents say that email is a waste of time, plus most people did not receive training in regard to email usage; this you could take as suggesting that content-based training should be provided ...

- Let us move through the data in a more structured way, exploiting the idea of the 3 main variable groups ...

- In doing so, we always have to check out the codebook to understand the meaning of the variables ...

Using the idea of data storytelling in layers:

- Set up layer 1 (a basic look at the data, starting from the rsearch question):

* Variable "q_5" captures negative aspects of email use (what does the value of "99" mean here?)
* Whereas variable "q_6" captures positive aspects ...

* When inspecting "q_5" (e.g. through a pie chart, bar chart, or histogram), we see that the answers "4" and "11" are most common; but what does this mean? We have to consult our codebook!
* We explore other variables, too, e.g. "q_7_score", "q_9_score", and "q_11" (this quantitative variable can have a very different interpretation, depending on the profession looked at)
* Which variables can provide more information? E.g. "q_17_score" mentions "waste of time" and "q_19a" plus "q_fpxariert" can be used to further specify what is behind this (lack of focus on content)

- Then, set up layer 2 (look at the training-related variables):

* Here, we should be looking for information how many people have actually received training in regard to email use ...
* And we will then try to relate this to the information provided by "q_5"

- Finally, we set up layer 3 (consider the third variable group, characteristics of the respondents):

* These variables tell us more on the composition of the sample ...
* What about age?
* What about "q2_score" (this is employment category)? This can help us to decide whether any issues with email usage are particularly pronounced in a specific employment category. If so, targeted training makes sense.

Hints for drafting the critique (part 2 of the assignment):

- The critique has the format of a report (several pages, not a poster).

- It is suggested to structure the critique such that it first outlines the principles of good visualization practice, that is, for example, the ideas of Gestalt Theory. Then, the data storytelling approach should be outlined (Nicole Nussbaumer Knaflic). This is the "theoretical" part of the critique.

- In terms of the relevant lessons in iLearn, the principles of good visualization practice are collected, in particular, in Lesson 4 (pages 6 and 10), then in Lesson 5 (page 15), and Lesson 6 (pages 4 and 5).

- After that, you will do the "practical" part of the critique, go through the poster, and consider and validate the title, the introduction (does it state the research question, the focus on education and email?), the data set description (correctly identifying the data as cross-sectional?) etc.

- You will consider the story told but, importantly, also the visuals and whether these conform to good practice e.g in terms of the choice of colour. And are the visuals appropriate for telling this story?

- As far as books to be used as references are concerned, you can use the texts suggested for the module in iLearn, such as the book by Andy Kirk, by Nicole Nussbaumer Knaflic, by Stephen Few, but also the work by William Cleveland. We also mention these in the module, from time to time.

- Apart from this, you are of course most welcome to use other literature, too.

Thanking you.

Good luck!

