FIT3152 Data analytics – 2025: Assignment 1

Your task	● Analyse the country level predictors of confidence in social organisations using data from the World Values Survey. ● This is an individual assignment.
Value	● This assignment is worth **25%** of your total marks for the unit. ● It has 40 marks in total.
Suggested Length	● 8 – 10 A4 pages, approximately 1,000 words (for your report) + extra pages as appendix (for your R script, clustering table, and report on how Generative AI used if required). ● Font size 11 or 12pt, single spacing.
Due Date	**11.55pm Wednesday 16th April 2025**
Submission	● Submit a single PDF file **and** single video file on Moodle. ● Note that submission of a video report is a hurdle requirement. ● Use the naming convention: *FirstnameSecondnameID.{pdf, mp4, mov etc.}* ● Turnitin will be used for similarity checking of all written submissions.
Generative AI Use	● In this assessment, you can use generative artificial intelligence (AI) in order to <u>search for R functions and examples to perform tasks that you specify</u> only. Any use of generative AI must be appropriately acknowledged (<u>see Learn HQ</u>). ● Note that a Generative AI statement is a hurdle requirement.
Late Penalties	● 5% (2 mark) deduction per calendar day for up to one week. ● Submissions more than 7 calendar days after the due date will receive a mark of zero (0) and no assessment feedback will be provided.

Instructions

Address each of the research questions below and report the results of your analysis and your interpretation of those results.

You are expected to include at least one high quality multivariate graphic summarising key results. You may also include other simpler graphs and tables. Report any assumptions you've made in your analysis. Include your R code as an appendix. Your R code must be machine readable text as the university requires all student submissions to be processed by plagiarism detection software. You must include a declaration on the use of Generative AI at the beginning of your written report (hurdle requirement) and if you used Generative AI, describe how it was used, as an appendix.

There are two options for compiling your written report:
(1) You can create your report using any word processor with your R code pasted in as machine-readable text as an appendix, and save as a pdf, or
(2) As an R Markdown document that contains the R code with results and discussion interleaved. Render this as a HTML file and save as a pdf.

Your video report should be less than 100MB in size. You may need to reduce the resolution of your original recording to achieve this. Use a standard file format such as .mp4, or mov for submission.

Software

It is expected that you will use R for your data analysis and graphics and tables. You are free to use any R packages you need but must document these in your report and include in your R code. You may use other software, such as Excel, to create the table of clustering data for Question 3(a).

Use of Generative AI

AI & Generative AI tools may be used in GUIDED ways within this assessment / task as per the guidelines provided.

In this assessment, you can use generative artificial intelligence (AI) in order to search for R functions and examples to perform tasks that you specify only. Any use of generative AI must be appropriately acknowledged (see Learn HQ).

If you do use Generative AI for your assignment, then you must include the statement "Generative AI was used in this assignment." in the introductory/first paragraph of your report. You must also include the following information as an appendix in your report: (1) the technology you used (e.g. ChatGPT), (2 the information that was generated (e.g. R code fragments), (3) the prompts used (i.e. the questions you asked), and (4) how the output was used in your work.

If you did not use generative AI in your assignment, then include the statement "Generative AI was not used in this assignment." in the introductory/first paragraph of your report.

Questions

The World Values Survey (WVS) is an international research program that studies the social, political, economic, religious and cultural values of people in the world. You can read more here: https://www.worldvaluessurvey.org/WVSContents.jsp

The aim of this assignment is to understand country-level differences in the predictors of confidence in social organisations reported by participants.

These social organisations cover aspects of society such as: religion, armed forces, the press, television, trade unions, police, the courts, government, political parties, parliament, public service, universities, elections, major companies, banks, and environmental organisations. They are indicated in your data by column names having the prefix "C". Predictor variables (attributes) include personal information such as age and gender, values such as trust in others, the importance of various things in life, feelings of security and attitudes towards the state and science.

Each student will be assigned a subset of organisations and predictor variables to study. Your task is to analyse the survey data overall, with a focus on the country you have been assigned. You may make use of any additional data you require to answer the following questions.

1. **Descriptive analysis. (4 Marks)**

   (a) Describe the data overall, including things such as dimension, data types, distribution of numerical attributes, variety of non-numerical (text) attributes, missing values, and anything else of interest or relevance.

2. **Focus country vs all other countries as a group. (14 Marks)**

   (a) Identify your focus country from the accompanying list (WVSFocusCountry.pdf). How do participant responses (attributes) for your focus country differ from the other countries in the survey(treating them as a group)?

   (b) How well do participant responses (attributes) predict confidence in social organisations in your focus country? Which attributes are the strongest predictors? Confidence in which social organisations can be more reliably predicted? Explain your reasoning.

   (c) Repeat Question 2(b) for the other countries as a group. How do these results compare to those of your focus country?

3. **Focus country vs cluster of similar countries. (12 Marks)**

   (a) Using a collection of social, economic, health, political or other indicators from external sources, identify at least 5 countries in in your survey data that are similar to your focus country using clustering. The references in this document list some data sources that may be relevant although you are encouraged to search more broadly. State the indicators used and describe how you calculated/identified similar countries. Copy and paste the table of values you used for your clustering into your report as an Appendix.

   (b) Repeat Question 2(b) for your cluster of countries. Comment on the similarity and/or difference between your results for this question and Question 2(c). That is, does the group of all other countries 2(c), or the cluster of similar countries 3(b) give a better match to the important attributes for predicting confidence in social organisations in your focus country? Explain why this might be the case.

4. **Video Presentation: (Submission Hurdle and 4 Marks)**

   Record a short presentation using your smartphone, Zoom, or similar method. Your presentation should be approximately 5 minutes in length and summarise your main findings for Sections 1 – 3, as well as describing how you conducted your research and any assumptions made. Pay particular emphasis to your results in Questions 2(c) and 3(b)

5 **Overall considerations (6 Marks)**

   This includes: the quality and clarity of your reasoning and assumptions; the strength of support for your findings; the quality of your writing in general and communication of results; the quality of your graphics throughout, including at least one high-quality multivariate graphic; the quality of your R coding.

Data

The data for this assignment is a reduced version of the World Values Survey Wave 7 data. The filename is "WVSExtract.csv". The data includes ordinal data coded on a numerical scale. For this assignment assume it is reasonable to treat these responses as numerical.

Create your individual data as follows:

```
rm(list = ls())
set.seed(12345678) # Your Student Number
VCData = read.csv("WVSExtract.csv")
VC = VCData[sample(1:nrow(VCData),50000, replace=FALSE),]
VC = VC[,c(1:6, sort(sample(7:46,17, replace = FALSE)), 47:53,
sort(sample(54:69,10, replace = FALSE)))]
```

Locate your focus country using the accompanying document FocusCountryByID.pdf. The document WVSCountryCodes.pdf identifies each country by name from its code.

Selected references and web links

World Values Survey Wave 7 (2017-2022)
https://www.worldvaluessurvey.org/WVSDocumentationWV7.jsp

The World Bank Data Collections (and Governance Indicators)
https://datacatalog.worldbank.org/collections
http://info.worldbank.org/governance/wgi/

Organisation for Economic Co-operation and Development (OECD)Data
https://data.oecd.org/

Global Health Security Index: Reports and Data
https://www.ghsindex.org/report-model/

World Health Organization
https://www.who.int/

Data fields and brief descriptor

Most fields are on integer scales over varying range. The convention is that larger numbers generally indicate greater agreement with statement or frequency of occurrence. Some exceptions given below. Fields in bold indicate confidence in social organisations.

You can access more detail on each field in the WVS-7 Master Questionnaire 2017-2020 English.pdf, linked from https://www.worldvaluessurvey.org/WVSDocumentationWV7.jsp

Column Name	Question/Brief Description
Country	Country of birth.
TPeople	Most people can be trusted.
TFamily	How much you trust your family.
TNeighbourhood	How much you trust your neighbourhood.
TKnow	How much you trust people you know personally.
TMeet	How much you trust people you meet for the first time.
VFamily	Importance in life: Family.
VFriends	Importance in life: Friends.
VLeisure	Importance in life: Leisure time.
VPolitics	Importance in life: Politics.
VWork	Importance in life: Work.
VReligion	Importance in life: Religion.
HOverall	Feeling of happiness overall.
HHealth	State of health overall.
HChoice	How much control do you have over your life?
HSatLife	How satisfied are you with your life?
HSatFin	How satisfied are you with the financial situation of your household?
HFood	In the last 12 months have you or your family: gone without enough food?
HCrime	In the last 12 months have you or your family: felt unsafe from crime?
HMedicine	In the last 12 months have you or your family: gone without medicine or medical treatment?
HIncome	In the last 12 months have you or your family: gone without cash income?
HShelter	In the last 12 months have you or your family: gone without safe shelter over your head?
EEquality	Income should be equal vs greater incentives for individual effort.
EPrivate	Private ownership of business and industry should be increased vs government ownership.
EGovernment	Government should take more responsibility for everyone vs people should take more responsibility for themselves.
ECompetition	Competition is good vs competition is harmful.
EHardWork	Hard work brings a better life vs it's more a matter of luck.
SSecure	How secure do you feel in the neighbourhood?
SJob	Worry about losing or not finding a job
SEducation	Worry about not being able to give one's children a good education
PIA	Which would you say is most important: 1. A high level of economic growth. 2. Making sure this country has strong defence forces. 3. People have more say about how things are done at their jobs and in their communities.

	4. Trying to make our cities and countryside more beautiful.
PIAB	Which would you say is next most important? See above for choices.
STBetter	Science and technology are making our lives healthier, easier, and more comfortable.
STOpportunity	Because of science and technology, there will be more opportunities for the next generation.
STFaith	We depend too much on science and not enough on faith.
STRight	One of the bad effects of science is that it breaks down people's ideas of right and wrong.
STImportant	It is not important for me to know about science in my daily life
STWorld	The world is better off because of science and technology
PNewspaper	Do you use the following information source (1) daily, (2) Weekly, (3) Monthly, (4) Less than monthly, (5) Never: Daily newspaper
PTelevision	See above. Information source: TV news
PRadio	See above. Information source: Radio news
PMobile	See above. Information source: Mobile phone
PEmail	See above. Information source: Email
PInternet	Information source: Internet
PSocial	See above. Information source: Social media (Facebook, Twitter, etc.)
PFriends	I See above. Information source: Talk with friends or colleagues
PDemImp	How important is it for you to live in a country that is governed democratically?
PDemCurrent	How democratically is this country being governed today?
PSatisfied	How satisfied are you with how the political system is functioning in your country these days?
MF	Respondent's sex (Male, Female)
Age	Age (in two digits).
Edu	Highest educational level: Respondent
Employment	Employment status: (1) Full time, (2) Part time, (3) Self employed, (4) Retired, (5) Spouse/not employed, (6) Student, (7) Unemployed, (8) Other.
CReligious	**Confidence in Religious Institutions (Church, mosque, temple, etc., whichever relevant)**
CArmedForces	**Confidence in the Armed Forces**
CPress	**Confidence in the press.**
CTelevision	**Confidence in television (companies).**
CUnions	**Confidence in labour (trade) unions.**
CPolice	**Confidence in the police.**
CCourts	**Confidence in the justice system/courts.**
CGovernment	**Confidence in the government.**
CPParties	**Confidence in political parties.**
CParliament	**Confidence in parliament.**
CCivilService	**Confidence in the civil (public) services.**
CUniversities	**Confidence in universities.**
CElections	**Confidence in elections.**
CMajCompanies	**Confidence in major companies.**
CBanks	**Confidence in banks.**
CEnvOrg	**Confidence in environmental protection movements.**